

VYSOKÁ ŠKOLA BÁŇSKÁ - TECHNICKÁ UNIVERZITA OSTRAVA
EKONOMICKÁ FAKULTA

KATEDRA SYSTÉMOVÉHO INŽENÝRSTVÍ

**Návrh řešení dashboardu pro prezentaci prodejních výsledků s využitím nástrojů
Microsoft Business Intelligence**

**The Proposal of a Dashboard Solution for Sales Results Presentation Utilizing Microsoft
Business Intelligence Tools**

Student: Bc. Jana Kaděrová
Vedoucí diplomové práce: Ing. Radek Němec

Ostrava 2012

Zadání diplomové práce

Student: **Bc. Jana Kaděrová**

Studijní program: N6209 Systémové inženýrství a informatika

Studijní obor: 6209T025 Systémové inženýrství a informatika

Téma: Návrh řešení dashboardu pro prezentaci prodejních výsledků s využitím nástrojů Microsoft Business Intelligence
The Proposal of a Dashboard Solution for Sales Results Presentation Utilizing Microsoft Business Intelligence Tools

Zásady pro vypracování:

1. Úvod
2. Teoretická východiska analýzy a návrhu aplikací Business Intelligence
3. Analýza výchozího stavu řešení prezentace prodejních výsledků
4. Návrh řešení dashboardu pro prezentaci prodejních výsledků
5. Závěr

Seznam použité literatury

Seznam zkratek

Prohlášení o využití výsledků diplomové práce

Seznam příloh

Přílohy

Seznam doporučené odborné literatury:

NOVOTNÝ, Ota a Jan POUR a David SLÁNSKÝ. *Business Intelligence: Jak využít bohatství ve vašich datech*. Praha: Grada Publishing, 2005. 256 s. ISBN 80-247-1094-3.

ECKERSON, Wayne W. *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. 2nd ed. New Jersey: John Wiley & Sons, 2011, p. 318. ISBN 978-0-470-58983-0.


LACKO, Ľuboslav. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby*. Brno: Computer Press, 2009. 456 s. ISBN 978-80-251-2887-9.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.


Vedoucí diplomové práce: **Ing. Radek Němec**

Datum zadání: 25.11.2011

Datum odevzdání: 27.04.2012


Ing. Eva Moravcová, CSc.
vedoucí katedry




prof. Dr. Ing. Dana Dluhošová
děkanka fakulty

Prohlašuji, že jsem celou práci, včetně všech příloh, vypracovala samostatně.

V Ostravě 27. dubna 2012

.....

PODĚKOVÁNÍ

Děkuji vedoucímu diplomové práce, panu Ing. Radku Němcovi, za odbornou pomoc a další cenné rady při zpracování mé diplomové práce. Také bych ráda poděkovala paní Ing. Jitce Andraškové za poskytnutí informací a za spolupráci.

V Ostravě 27. dubna 2012

.....

Obsah

1. Úvod.....	4
2. Teoretická východiska analýzy a návrhu aplikací Business Intelligence	5
2.1. Business Intelligence	5
2.2. Data, informace, poznatky a znalosti.....	7
2.3. Vrstvy a komponenty BI.....	8
2.4. Multidimenzionalita - princip ukládání dat v BI	21
2.5. Dimenzionální modelování.....	23
2.6. Dashboard - vizuální zobrazení BI řešení	27
2.7. Shrnutí	31
3. Analýza výchozí stavu řešení prezentace prodejních výsledků	32
3.1. Představení společnosti.....	32
3.2. Výchozí stav řešení prezentace prodejních výsledků	32
4. Návrh řešení dashboardu pro prezentaci prodejních výsledků	34
4.1. Sběr požadavků	34
4.2. Analýza uživatelských požadavků.....	35
4.3. Návrh, modelování a implementace řešení	37
4.4. Nasazení řešení do provozu	51
4.5. Shrnutí	54
5. Závěr	55
Seznam použité literatury.....	56
Seznam zkratk.....	57

1. Úvod

Business Intelligence se v současné době stává velice populární a v mnoha případech nezastupitelnou součástí chodu společností ve všech odvětvích podnikání. Uživatelé společnosti, se kterou spolupracuji v rámci této diplomové práce, v současné době nevyužívají veškeré funkcionality, která je v rámci Business Intelligence k dispozici.

Cílem práce je proto navrhnout řešení pro prezentaci prodejních výsledků formou dashboardu prostřednictvím nástrojů Microsoft Business Intelligence. Součástí návrhu bude také specifikace způsobu čerpání dat ze zdrojových systémů (ETL proces) a způsobu jejich uložení pro potřeby efektivní vizualizace pomocí navrhovaného dashboardu.

V práci bude využito metody skupinového rozhovoru pro sběr požadavků na navrhovaný dashboard a dále pak bude využito dimenzionálního modelování pro návrh datového modelu datového skladu, nad kterým bude vytvořen výsledný dashboard.

První částí práce bude čtenáři přiblížena základní teorie a terminologie užívaná v oblasti Business Intelligence, která bude také použita v dalších částech práce.

Druhá část práce bude zaměřena na přiblížení výchozího stavu řešení prodejních výsledků, kde bude popsána současná podoba řešení.

V následující části budou specifikovány jednotlivé kroky, které budou nezbytné při návrhu řešení dashboardu pro prezentaci prodejních výsledků.

Závěrem budou shrnuty výsledky práce a dále budou uvedeny náměty k dalšímu řešení.

2. Teoretická východiska analýzy a návrhu aplikací Business Intelligence

2.1. Business Intelligence

2.1.1. Pojem Business Intelligence

V době, kdy konkurence na rychle se měnícím trhu, potřeby zákazníků a nabídky firem se zvyšují, je důležité, aby přijímání rozhodnutí bylo téměř okamžité. Rychlé a kvalitní rozhodnutí je založeno na kvalitních informacích, které jsou k dispozici ve chvíli, kdy jsou potřeba a jsou uloženy na správném místě a v požadované podobě. Aby toto bylo zajištěno, je nezbytné fungování procesů, jejichž účelem je poskytování informací pro podporu rozhodování.

Novotný, Pour a Slánský (2005) definují pojem Business Intelligence jako sadu procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy podniků a organizací a podporovat tak analytické a plánovací činnosti, které jsou postaveny na principech multidimezionálních pohledů na podniková data.

2.1.2. Vývoj BI

Pojem Business Intelligence poprvé definoval Howard Dresner ze společnosti Gartner Group teprve v roce 1989. Kořeny konceptů a technologií však můžeme nalézt již v systémech MIS (Management Information Systems, Manažerské informační systémy) v 70. letech 20. století, jejichž vstupy byly data z transakčních systémů. Řídící pracovníci tak měli určité komplexní přehledy a sestavy agregovány podle různých hledisek. Nevýhodou však bylo to, že přehledy či sestavy se dostaly k manažerům až s určitou časovou prodlevou, která vznikala v důsledku toho, že požadavky byly nejprve odeslány vývojovému týmu MIS, který je teprve na základě nich vytvořil.

Operativnější výsledky byly poskytovány systémy DSS (Decision Support Systems, Systémy pro podporu rozhodování). Ty, na rozdíl od systémů MIS, byly nasazovány na úrovni taktického a strategického rozhodování, takže řídicím pracovníkům například poskytovaly výsledky poměrně složitých analýz.

V 80. letech 20. století se objevily informační systémy pro vrcholové řízení (EIS, Executive Information Systems), které představovaly strategickou úroveň rozhodování. Tyto systémy se již vyznačovaly tím, že dokázaly zpracovat i velké množství dat. Součástí systémů byly také multidimenzionální pohledy a možnost drill down v datech, reportovací služby

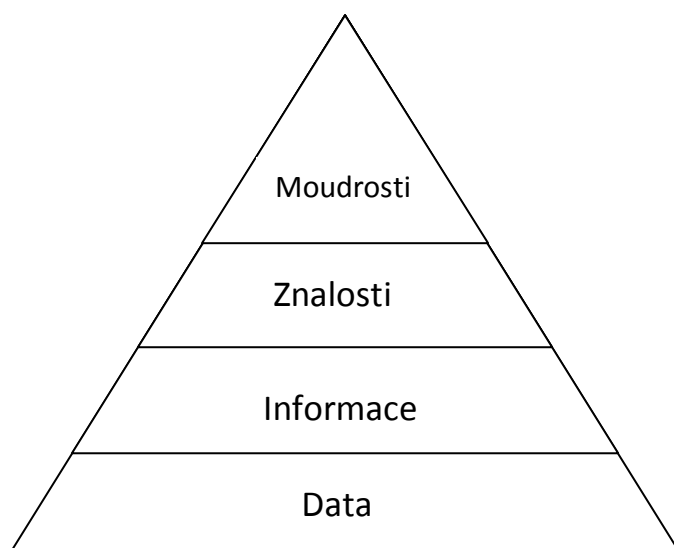
a analýzy trendů, což poskytovalo uživatelům mnohem dynamičtější přístup k práci s daty. Jednotlivé prvky současných Business Intelligence systému se tak v podstatě transformovaly z konceptu systémů EIS.

V posledních letech se stále častěji můžeme setkat s umělou inteligencí v Business Intelligence systémech (propojením Business Intelligence systému s expertními systémy) a dále také s přesunem systémů na internet (Cloud computing) a v neposlední řadě s vývojem Business Intelligence systémů pro mobilní zařízení.

2.2. Data, informace, poznatky a znalosti

Mezi základní prvky Business Intelligence patří data, informace, poznatky a znalosti. Proces transformace dat na informace a převod těchto informací na poznatky prostřednictvím objevování nazýváme Business Intelligence, jak tvrdí Lacko (2009, s. 14) .

„Přeměnu dat na informace, informací na znalosti a budování „moudrosti“ na základě znalostí můžeme zobrazit na hierarchické pyramidě informačních úrovní. Základem všeho jsou data. Data obsahují jednoduchá fakta, přičemž se samozřejmě předpokládá, že někde uvnitř množiny dat jsou ukryty určité informace. Tyto informace však vystoupí na povrch až tehdy, když k datům přidáme souvislosti. Když do hry vstoupí kromě informací i tvořivá inteligence, získáme znalosti. Pokud tyto znalosti zobecníme, získáme „moudrost“, to znamená schopnost přesného zhodnocení znalostí a jejich následné uplatnění v reálné praxi.“
[Lacko, 2009, s. 14]



Obr. 2. 1 - Hierarchická pyramida informačních úrovní

Zdroj: [Lacko, 2009, s.15 - vlastní přepracování]

2.3. Vrstvy a komponenty BI

2.3.1. Vrstvy BI

Různorodost řešení systémů Business Intelligence, stejně jako různorodost Business Intelligence nástrojů, vedla k tomu, že obecná koncepce architektury řešení se rozdělila do několika vývojových větví, čímž se konkrétní aplikace řešící reálné problémy od sebe liší.

Během vývoje oblasti se však ustálila obecná koncepce architektury řešení Business Intelligence, kde lze identifikovat tyto vrstvy zajišťující:

a. Extrakci, transformaci, čištění a nahrávání dat

Vrstva pro extrakci, transformaci, čištění a nahrávání dat zabezpečuje sběr a následný přenos dat ze zdrojových systémů do vrstvy ukládající data do Business Intelligence systému.

Hlavní komponenty této vrstvy jsou ETL systémy, neboli systémy pro extrakci, transformaci a přenos dat a systémy EAI, což jsou systémy pro integraci aplikací.

b. Ukládání dat

Tato vrstva zajišťuje ukládání dat do databázových komponent řešení Business Intelligence a následnou aktualizaci a správu těchto dat.

Mezi databázové komponenty Business Intelligence patří zejména datové sklady a dále pak dočasná uložení dat, datová tržiště a operativní uložení dat.

c. Analýzu dat

Vrstva pro analýzu dat pokrývá činnosti spojené s vlastním zpřístupněním dat a jejich analýzou.

Mezi analytické komponenty můžeme zařadit systémy OLAP, které jsou zaměřené na pokročilé a dynamické analytické úlohy a dále pak proces Dolování dat, který provádí sofistikovanou analýzu velkého množství dat.

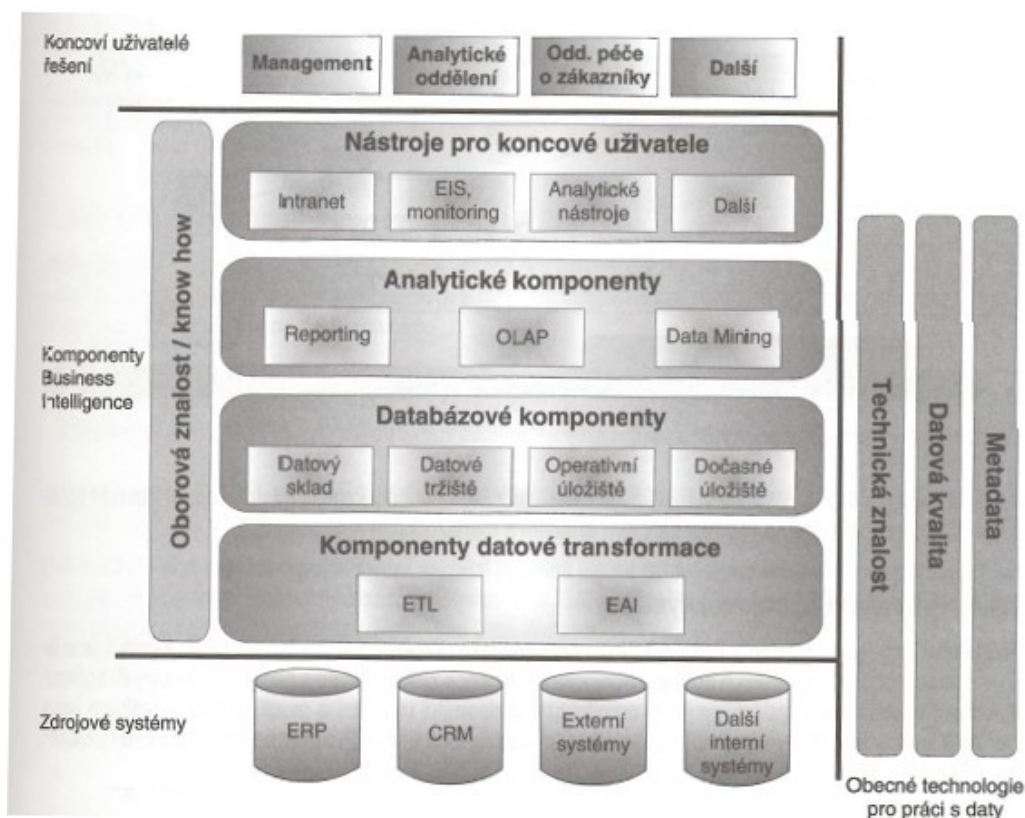
d. Prezenci dat

Prezentační vrstva zprostředkovává komunikaci mezi koncovými uživateli a ostatními komponentami Business Intelligence systému. Jedná se především o sběr uživatelských požadavků na analytické operace a následně pak prezenci výsledků těchto analýz.

Portálové aplikace založené na webovských technologiích, systémy EIS a různé další analytické aplikace patří mezi hlavní nástroje koncového uživatele.

e. Vrstva oborové znalosti

Do této vrstvy patří oborová znalost, neboli know-how a tzv. best-practices popisující konkrétní situaci nasazování řešení Business Intelligence v organizaci.

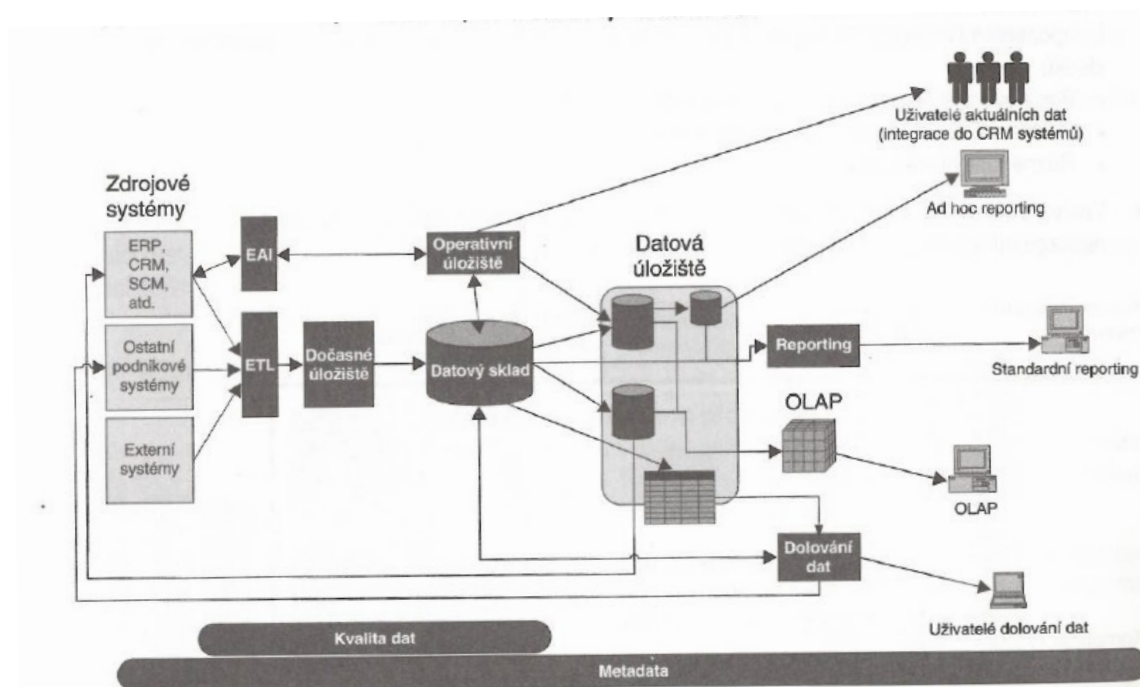


Obr. 2. 2 - Obecná koncepce architektury Business Intelligence

Zdroj: [Novotný, Pour, Slánský, 2005, s. 27]

2.3.2. Komponenty řešení BI

Na začátek je třeba upozornit na to, že v závislosti na rozsahu řešení systému Business Intelligence a podle potřeb koncových uživatelů se může výrazně měnit uspořádání jednotlivých komponent.



Obr. 2. 3 – Hlavní komponenty Business Intelligence a jejich vazby

Zdroj: [Novotný, Pour a Slánský, 2005, s. 28]

a. *Produkční (zdrojové) systémy*

Novotný, Pour a Slánský (2005) definují produkční (zdrojové) systémy (také někdy označované jako primární, transakční, OLTP či „legacy“) jako takové systémy podniku, ze kterých aplikace Business Intelligence získávají data a nepatří do skupiny BI aplikací. Vlastnosti všech těchto systémů je jejich architektura podporující ukládání a modifikaci dat v reálném čase. Oproti BI aplikacím tyto systémy nejsou navrženy pro analytické úlohy. Příkladem mohou být ERP¹, SCM², CRM³ systémy, specializované systémy pro podporu personálních oddělení, pro podporu finančních oddělení a další. Zdrojem pro řešení BI nemusí být pouze vnitřní systémy podniku, ale i externí systémy (např. databáze podnikatelských subjektů, telefonní seznamy, výstupy statistických úřadů či vládních institucí apod.).

¹ Enterprise Resource Planning

² Supply Chain Management

³ Customer Relationship Management

„Produkční systémy jsou hlavním, a často i jediným vstupem do BI. V praxi je většinou spektrum zdrojových, resp. produkčních systémů pro BI velmi různorodé a heterogenní jak obsahově, tak technologicky. Úkolem řešení BI je pak zajistit analýzu těchto zdrojů z pohledu potřeb řízení firmy, výběr relevantních dat pro řízení, a následně jejich vzájemnou integraci. Právě tato část projektů BI je pracovně, časově i finančně nejnáročnější, ale představuje zcela nezbytný předpoklad úspěšných aplikací Business Intelligence“, jak tvrdí Novotný, Pour a Slánský (2005, s. 29).

b. ETL - nástroje pro výběr, transformaci a přenos dat

Jednou z nejvýznamnějších komponent celého Business Intelligence je ETL, často také označováno jako datová pumpa. Jejím úkolem je zajistit převod dat ze zdrojových datových struktur do cílových struktur datového uložiště. Tento proces probíhá ve třech krocích:

- Extrakce (Extraction) - kdy jsou data získávána a vybírána ze spektra zdrojových datových systémů nejrůznějšího charakteru,
- Transformace (Transformation) - data jsou upravena do požadované formy a vyčištěna, tzn. je odstraněna nekonzistence, odstraní či změně se nekvalitní data apod.,
- Nahrání (Load) - vložení dat do specifických datových struktur, resp. datových schémat, datového skladu.

Nástroje ETL pracují v dávkovém (batch) režimu, tzn. data jsou přenášena v opakujících se časových intervalech (denně, týdně, měsíčně) a obvykle v hodinách, kdy jsou zdrojové systémy nejméně vytíženy a nehrozí tak kolaps.

Jak uvádí Tvrdíková (2008) trendem v oblasti ETL nástrojů je slučování s nástroji pro správu metadat a s nástroji pro zajištění datové kvality a dále to, že se ETL nástroje stávají jednou z běžných součástí standardních databázových nástrojů.

c. Enterprise Application Integration - integrace podnikových aplikací

Nástroje pro integraci podnikových aplikací (EAI) jsou ve většině případů využívány ve vrstvě zdrojových systémů jako komplex nástrojů, které integrují primární podnikové systémy, čímž se razantně redukuje počet vzájemných rozhraní. Dle Novotného, Poura a Slánského (2005) pracují nástroje EAI na dvou úrovních:

- úroveň datové integrace, kde jsou EAI platformy využity pro integraci a distribuci dat,
- úroveň aplikační integrace, kde jsou EAI platformy použity pro sdílení určitých vybraných funkcí informačních systémů.

EAI platformy pracují na rozdíl od ETL nástrojů v reálném čase. To přináší využití datové integrace nástrojů EAI pro přenos dat do datových úložišť v reálném čase, čímž vzniká nová generace datových skladů, tzv. Real-Time Data Warehouse.

d. Data Staging Area - dočasná uložení dat

Dočasná uložení dat (DSA) slouží k prvotnímu ukládání netransformovaných dat ze zdrojových systémů před jejich vlastním zpracováním. Jejich cílem je urychlit výběr dat a uplatnění tak nacházejí:

- u produkčních systémů, které jsou neustále zatíženy a je tak nutné přenášet jejich data s minimálním dopadem na jejich výkon,
- u systémů pracujících např. s textovými soubory apod., kde je potřeba data před zpracováním nejprve převést do databázového formátu.

Jak uvádí Tvrdíková (2008), obsah DSA tvoří data s následujícími charakteristikami:

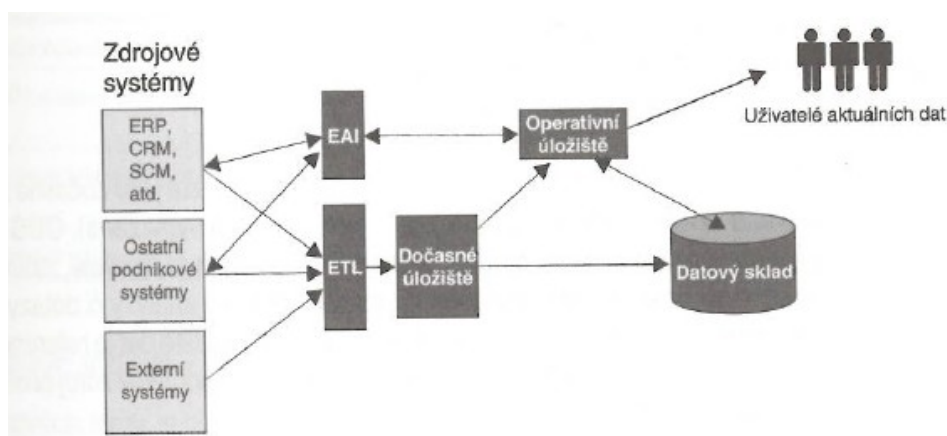
- detailní - data nejsou agregována,
- nekonzistentní - data nejsou kontrolována s externími číselníky či s ostatními daty v datovém skladu,
- pouze aktuální - přenášejí se pouze aktuální data ze zdrojových systémů, která neobsahují historii,
- měnící se - při každém výběru se berou pouze data, která ještě nebyla zpracována a po jejich zpracování a přenosu do dalších komponent BI se z DSA odstraní,
- ve stejné struktuře - data v DSA jsou v přesně stejné struktuře jako ve zdrojových systémech.

e. Operational Data Store - operativní uložení dat

Operativní uložení dat (ODS) vznikla na základě potřeby zpracovat klíčová data v relativně krátké době a učinit tak operativní rozhodnutí. Jedná se však o komponentu, kterou nemusíme nalézt ve všech řešeních Business Intelligence.

Novotný, Pour a Slánský (2005) tvrdí, že existují dva základní přístupy k návrhu operativního uložení dat:

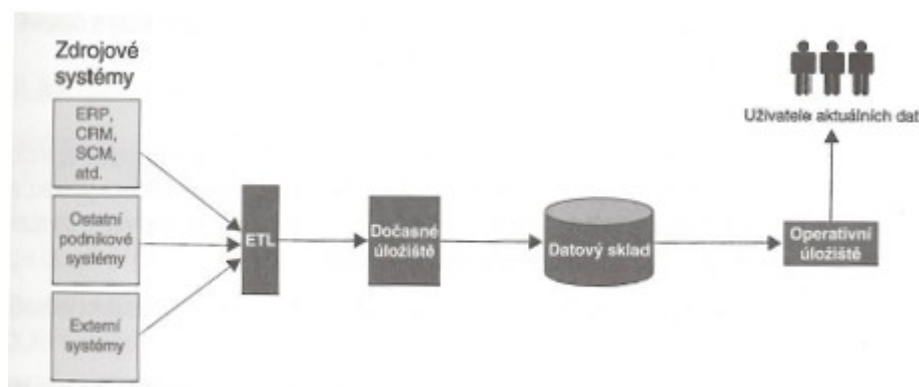
- První přístup definuje ODS jako jednotné místo datové integrace aktuálních dat ze zdrojových systémů. Jedná se tedy o zdroj pro sledování konsolidovaných agregovaných dat téměř v reálném čase, tedy s minimální dobou odezvy pro zpracování. ODS pak v mnoha případech slouží jako centrální databáze základních číselníků nebo jako podpora interaktivní komunikace se zákazníky, např. v call centrech, kde dotyčný pracovník dostává aktuální konsolidovaná data o zákazníkovi, jeho profil, seznam použitých produktů apod. Při tomto přístupu pracuje operativní uložení dat s nástroji EAI, kdy dochází k obousměrnému přenosu aktuálních dat z a do zdrojových systémů v reálném čase. Navíc mohou být data v ODS doplněna z datového skladu nebo naopak datový sklad může být doplněn aktuálními konsolidovanými daty z ODS.



Obr. 2. 4 – ODS jako koncept jednotného místa datové integrace aktuálních dat ze zdrojových systémů

Zdroj: [Novotný, Pour a Slánský, 2005, s. 31]

- V druhém přístupu pracuje operativní uložení dat, pouze s daty, která pocházejí z datového skladu. Je tak v podstatě podmnožinou již existujícího datového skladu, ale na rozdíl od něj obsahuje pouze aktuální část dat a nikoliv celou historii. Cílem takto navrženého ODS je podporovat relativně jednoduché dotazy nad malým množstvím aktuálních analytických dat.



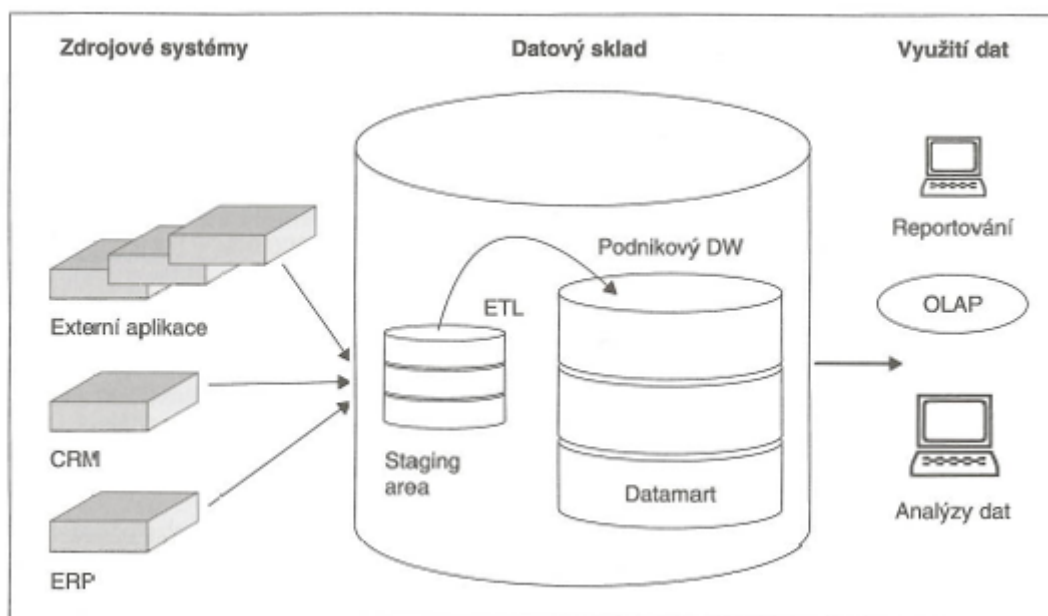
Obr. 2. 5 – ODS jako koncept databáze aktuálních dat odvozené z datového skladu

Zdroj: [Novotný, Pour a Slánský, 2005, s. 31]

Největším rozdílem mezi dočasným úložištěm dat a operativním úložištěm dat je v jejich použití. Zatímco cílem DSA je dočasné uložení dat před jejich zpracováním v datovém skladu (přičemž po zpracování jsou tato data smazána), ODS tvoří databázi podporující analytické procesy. Z toho vyplývá, že do DSA koncoví uživatelé nemají přístup na rozdíl od ODS, která je vytvářena právě proto, aby byly zpřístupněny data pro analýzy či dotazy s minimálním zpožděním oproti jejich pořízení.

f. Data Warehouse - datový sklad

Datový sklad (DWH) popisuje Tvrdíková (2008) jako ucelenou databázi, optimalizovanou pro dotazování a analýzu dat, společně s nástroji, které dotazy, analýzy a kvalitní prezentaci výstupů umožňují.



Obr. 2. 6 – Architektura datového skladu

Zdroj: [Tvrdíková, 2008, s. 109]

Definice datových skladů existuje mnoho. Nejpoužívanější definicí je však definice jednoho ze zakladatelů Data Warehousingu Billa Inmona (2005), ten definuje datový sklad jako „integrovaný, subjektově orientovaný, stálý a časově rozlišitelný souhrn dat, uspořádaný pro podporu potřeb managementu.“

Význam pojmů lze pak interpretovat takto:

- Integrovaný - data jsou ukládána v rámci celého podniku, ne pouze v rámci oddělení,
- Subjektově orientovaný - data jsou členěna podle typu, ne podle aplikací, ve kterých vznikla,
- Stálý - datové sklady jsou koncipovány pouze pro čtení, což znamená, že do datových skladů nelze uživatelskými nástroji data vkládat nebo je měnit a existují zde po celou dobu života datového skladu,

- Časově rozlišitelný - data uvnitř DWH musí sebou nést informaci o dimenzi času, aby bylo možné provádět analýzy za určitá období včetně historických.

Cílem datových skladů je poskytnout čitelné a organizované informace v reálném čase z podnikových informačních systémů či externích zdrojů, které obsahují velké množství hodnotných dat využitelných při řízení společnosti.

Rozdíl mezi daty, která jsou uložena v datovém skladu, a daty z produkčních databází shrnuje následující tabulka:

	Datový sklad	Produkční databáze
Procesy	Dotazy koncových uživatelů, OLAP	Vstup dat, dávkové zpracování, OLTP
Funkce	Podpora rozhodování	Zpracování dat, podpora podnikových operací
Data	Předmětně orientována, aktuální i historická, sumarizovaná	Procesně orientována, aktuální hodnoty, detailní
Užití	Ad hoc, částečně opakující se reporty a strukturované aplikace	Strukturované, opakované

Tab. 2. 1 - porovnání základních charakteristik produkčních databází a datových skladů

Zdroj: [Novotný, Pour a Slánský, 2005, s. 33 - vlastní přepracování]

V současné době představuje technologie datových skladů jeden z nejvýznamnějších trendů v rozvoji podnikových informačních systémů.

g. Data Mart - datové tržiště

Datová tržiště (DMA) pracují na obdobném principu jako datové sklady, ale na rozdíl od nich jsou určeny pro omezený okruh uživatelů. Jedná se například jednotlivé oddělení, divize, pobočky společnosti apod. Dalo by se říct, že se jedná o problémově orientovaný datový sklad vytvořený k řešení a analýze konkrétní problematiky daného okruhu uživatelů.

Podstatou datových tržišť jsou tak decentralizované datové sklady, které se postupně integrují do celopodnikového řešení Business Intelligence. Mohou ale také sloužit jako mezistupěň při transformacích dat z produkčních databází do celopodnikového datového skladu.

Výsledkem vytváření datových tržišť je zkrácení doby návratnosti investic, snížení nákladů a podstatné zmenšení rizika při jejich zavádění, jak uvádí Novotný, Pour a Slánský (2005).

h. OLAP databáze

Aplikace Business Intelligence využívají OLAP (On-Line Analytical Processing) databáze, které umožňují uživateli provádět efektivní vícekritériální analýzy a ad-hoc analýzy s přijatelnou rychlostí.

OLAP databáze jsou tvořeny jednou nebo více souvisejících vícerozměrných OLAP kostek, které již zahrnují předzpracované agregace dat podle definovaných hierarchických struktur jednotlivých dimenzí.

Jak uvádí Tvrdíková (2008), podle typu ukládání a způsobu zpracování dat se OLAP technologie dělí do čtyř skupin:

- ROLAP (Relational OLAP) reprezentuje přímý přístup k datům relačního systému, což znamená, že pro uložení dat se používají standardní relační databáze a data z nich jsou vybírána pomocí SQL dotazů,
- MOLAP (Multidimensional OLAP) využívají specifickou multidimenzionální databázi. Informace jsou navrženy jako množina multidimenzionálních matic, tzv. OLAP kostkách, a jsou aktualizovány a doplňovány v určitém pravidelném intervalu.

- HOLAP (Hybrid OLAP) je kombinací předchozích dvou architektur. Detailní data jsou uložena v relační databázi a agregované hodnoty jsou uloženy v OLAP kostkách.
- DOLAP (Desktop OLAP) je nejmladší architektura OLAP databází, která umožňuje připojení k centrálnímu úložišti OLAP dat, kde se stáhne potřebná podmnožina kostky na lokální počítač. Všechny analytické operace pak probíhají nad touto lokální kostkou, takže uživatel nemusí být připojen k serveru. To je výhodné především pro podporu mobilních aplikací.

i. Data mining - dolování dat

Dolování dat je analytická technika, která za pomoci speciálních algoritmů umožňuje automaticky objevovat strategické informace v uložených datech.

Proces dolování dat lze charakterizovat jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází - datového skladu, jak uvádí Novotný, Pour a Slánský (2005).

Dolování dat je založeno na množství matematických a statistických technik. Důležitou vlastností je, že se nejedná o předem specifikované analýzy, ale o analýzy odvozené z obsahu dat. Jedná se například o genetické algoritmy, rozhodovací stromy, neuronové sítě, clustering apod.

Cílem úloh pro dolování dat je, obdobně jako u většiny úloh Business Intelligence, poskytnout strategické informace širokému spektru manažerů ve společnosti.

Aplikace Business Intelligence kromě výše uvedených komponent využívají i tzv. obecné komponenty, jejichž úkolem je správa a manipulace s daty. Jedná se například o:

a. Nástroje pro zajištění datové kvality

Datová kvalita je jednou ze základních a nezbytných vlastností datových skladů a operativních úložišť, díky faktu, že pro úspěch nasazení řešení Business Intelligence je třeba korektní obsah zachycující reálnou situaci společnosti. Docházelo totiž k tomu, že data, která byla zpracována z různých zdrojů, v mnoha případech neodpovídala realitě nebo obsahovala konfliktní hodnoty.

Jak definuje Tvrdíková (2008, s. 117) „kvalitní data jsou taková data, která odpovídají realitě, jsou úplná a konzistentní.“

Cílem těchto nástrojů je zpracovat data tak, aby byla zajištěna jejich:

- úplnost - jsou identifikována a ošetřena data, která chybí nebo jsou z různých důvodů nepoužitelná
- soulad - všechna data, která nejsou ve standardním formátu, jsou identifikována a ošetřena
- konzistence - jsou identifikována a ošetřena data, která obsahují hodnoty, jež reprezentují konfliktní informace
- přesnost - jsou identifikována a ošetřena data, která jsou zastaralá nebo jsou nepřesná
- unikátnost - pokud existují záznamy, které jsou duplicitní, jsou identifikovány a ošetřeny
- integrita - jsou identifikována a ošetřena všechna data, která neobsahují veškeré definované vztahy vůči ostatním datům.

b. Nástroje pro správu metadat

Metadata jsou definována jako data o datech a dodávají kontext a smysl jinak nepřesně popsanému shluku informací. Nástroje pro správu metadat jsou hlavním prostředkem pro zjištění obsahu a stavu informačních systémů a umožňují tak snadnější pochopení jejich principů a funkcionality.

Metadata mohou být rozdělena do dvou skupin:

- věcná metadata - jsou tvořena informacemi o věcném obsahu řešení, tzn. je přidán kontext a význam jednotlivým hodnotám,
- technická metadata - obsahují informace o konfiguraci jednotlivých informačních systémů a souvisejících technických procesů.

Nástroje pro správu metadat společně s nástroji pro zajištění kvality dat umožňují pořizovat, spravovat, sdílet a poskytovat bezchybná a přesná data.

*c. **Oborová znalost - Know-how***

Oborová znalost definuje a formuje podobu celkové řešení Business Intelligence, na rozdíl od ostatních komponent, které jsou pouze nástroji k jeho vytvoření.

Oborovou znalost lze definovat jako:

- znalost fungování prostředí, kde se řešení BI implementuje,
- znalost možností technologií BI současně se znalostí nejvhodnějších řešení pro danou problematiku, tzv. Best practices.

Oborová znalost tak tvoří nezbytnou komponentu každého řešení Business Intelligence.

*d. **Technická znalost***

Na rozdíl od oborové znalosti, je technická znalost znalostí konkrétních technologií použitých během implementace řešení Business Intelligence. Jedná se tak například o znalost z problematiky návrhu HW architektury, programovacích jazyků, znalost databází, OLAP technologií a mnoha dalších.

2.4. Multidimensionalita - princip ukládání dat v BI

Data, která jsou uložena v provozních systémech, nejsou obvykle vhodná pro analytické účely, jelikož jsou primárně určena pro transakční zpracování (OLTP), kde je zapotřebí jejich normalizace. Pro analýzu a prezentaci v aplikacích Business Intelligence je zapotřebí, aby byla data uložena v odlišné struktuře, které umožňuje jejich rychlé čtení. Nejefektivnějším principem se jeví tzv. princip multidimensionality, který můžeme chápat jako uspořádání dat v logických celcích dle účelu jejich použití.

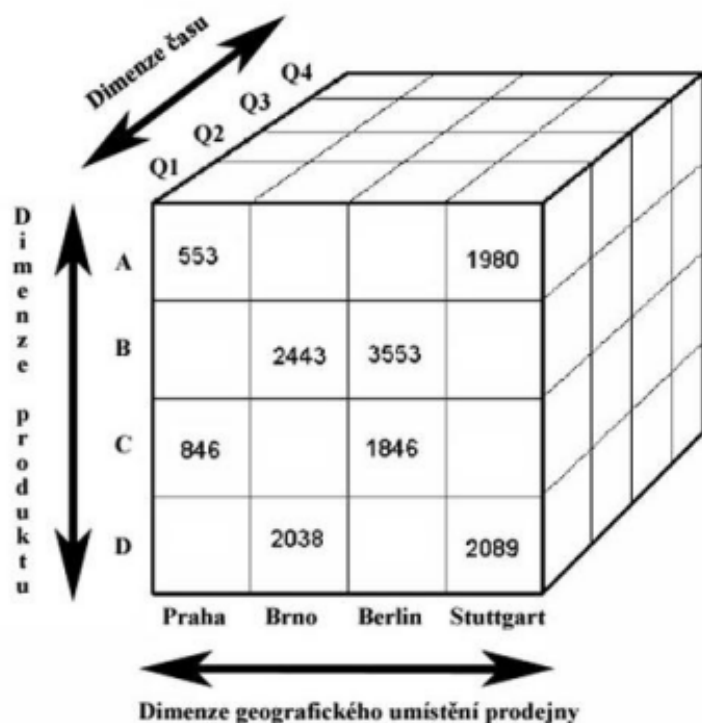
Multidimensionálně uspořádané data mohou být uložena dvěma způsoby:

- formou relační databáze,
- formou multidimensionální (OLAP) databáze.

Principem ukládání multidimensionálních dat v relační databázi je propojení tabulek dimenzí s tabulkami faktů. Naopak v multidimensionální (OLAP) databázi jsou data ukládána ve speciální logické datové struktuře. Obě formy uložení multidimensionálních dat však mají podobu tzv. datové kostky.

2.4.1. Datová kostka (Data Cube)

Datové kostky jsou vícerozměrným (vícedimenzionálním) rozšířením databázových tabulek, přičemž počet dimenzí může být teoreticky nekonečný (záleží na konkrétní implementaci).



Obr. 2. 7 - ukázka datové kostky

[Zdroj: <http://czm.fel.cvut.cz>]

„Datovou kostku si lze představit jako datovou strukturu OLAP databáze, která obsahuje hodnoty ukazatelů získané z jediné tabulky faktů a agregované podle hierarchie definovaných dimenzí.“ Novotný, Pour a Slánský (2005, str. 180).

Ukazatele z tabulky faktů reprezentují jednotlivé buňky kostky. Hrany kostky zase reprezentují jednotlivé dimenze.

2.5. Dimenzionální modelování

Jelikož datové sklady obsahují velký objem dat, vzniká problém s výkonem a rychlostí prováděných dotazů. Dimenzionální modelování nám poskytuje možnost, jak zvýšit výkon jednotlivých dotazů, aniž by byla ovlivněna integrita dat. Principem je návrh tabulek dimenzí, tabulek faktů, identifikace jednotlivých metrik, neboli ukazatelů, a dále pak návrh hierarchií uvnitř tabulek dimenzí.

2.5.1. Tabulka faktů

Tabulku faktů definuje Kimball (2002) jako elementární tabulku dimenzionálního modelu, ve které jsou uloženy číselné výkonnostní ukazatele z činnosti, kterou se společnost zabývá. Jedná se např. o objem prodeje, objem vyprodukovaného artiklu apod. Tyto ukazatele jsou v podstatě dílčí části dat ze zdrojových systémů, které mohou být dle potřeby předsumarizované. Úroveň detailnosti, ve které budou data ukládána v datovém skladu, vymezuje tzv. granularita. Obecně jde o detailnost popisu jednotlivých transakcí.

Charakteristiky typické pro tabulku faktů:

- každá tabulka faktů obsahuje cizí klíče z tabulek dimenzí,
- tabulka faktů obsahuje obvykle malé množství sloupců,
- v porovnání s tabulkou dimenzí, obsahuje tabulka faktů velký počet řádků.

2.5.2. Tabulky dimenzí

Tabulky dimenzí jsou nedílnou součástí dimenzionálního modelu. Obsahují atributy popisující detaily k jednotlivým řádkům tabulky faktů. Tím je zajištěno snadnější porozumění nutné k analýze dat.

Charakteristiky typické pro tabulky dimenzí:

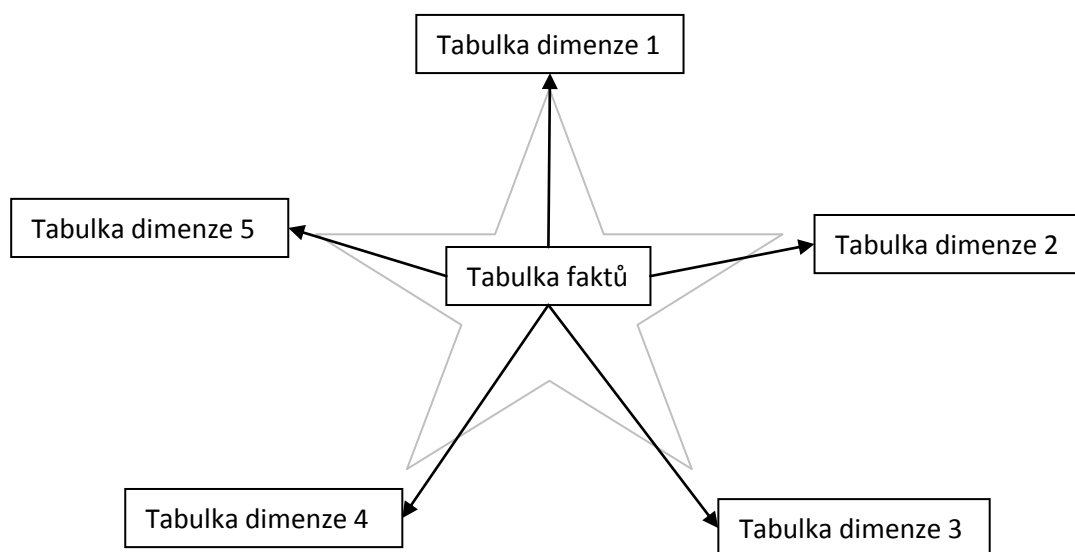
- tabulky dimenzí obsahují popisné informace k číselným hodnotám v tabulce faktů,
- tabulky dimenzí jsou obvykle denormalizované, což způsobuje, že obsahují velké množství sloupců,
- tabulky dimenzí obsahují výrazně méně řádků v porovnání s tabulkou faktů,
- atributy tabulky dimenzí se typicky používají pro záhlaví řádků a sloupců v reportech.

V tabulkách dimenzí se mohou stanovit hierarchie v závislosti na tom, jak na sebe hodnoty logicky, či hierarchicky navazují. Typickým příkladem je hierarchie v dimenzi času: minuta - hodina - den - týden - měsíc - rok.

Datový model, který je vytvořen na základě principů dimenzionálního modelu, by měl mít jednu ze dvou základních podob multidimenzionálního schématu.

a. Star schéma - schéma hvězdy

Schéma hvězdy obsahuje pouze jednu tabulku faktů a několik tabulek dimenzí, které jsou denormalizovány. Tak je docíleno vyšší rychlosti oproti schématu sněhové vločky, neboť je zredukována nutnost spojovat podřízené tabulky dimenzí. Nevýhodou může být rozsáhlost dimenzí, která stojí za nemožností procházet tabulky na vysoké úrovni detailu.

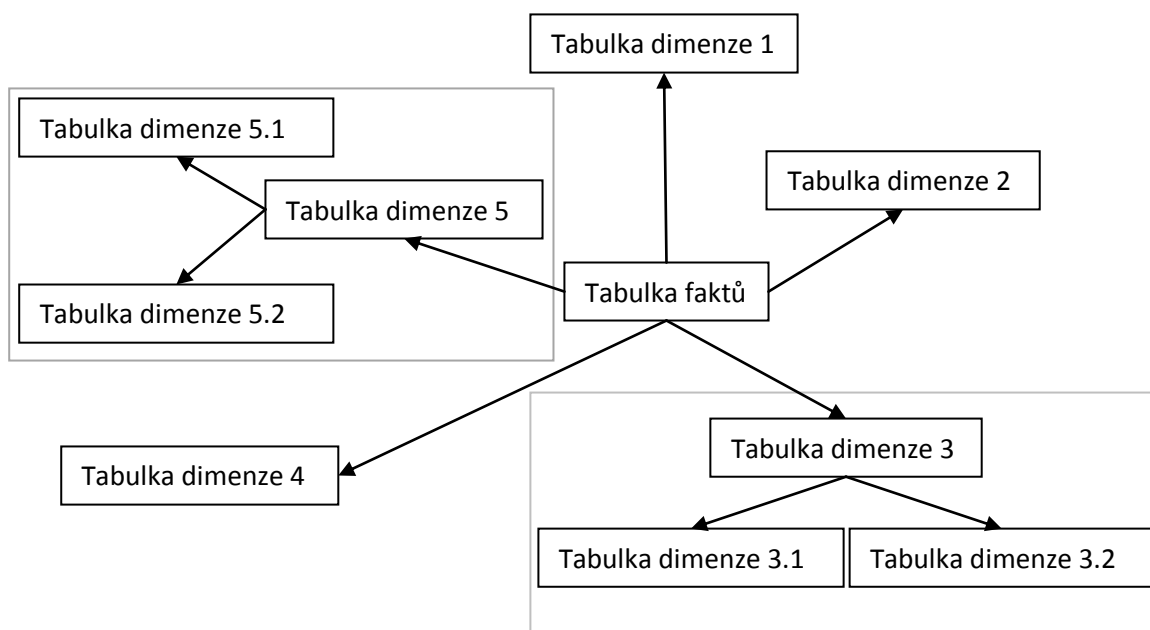


Obr. 2. 8 - Star (hvězdicové) schéma

[Zdroj: vlastní]

b. Snowflake schéma - schéma sněhové vločky

Schéma snowflake je v podstatě výsledkem normalizace tabulek dimenzí ve star schématu. Je tak docíleno možnosti dostat se snadno na vysokou úroveň detailu bez nutnosti vytváření velkých tabulek dimenzí. K tomu je ale zapotřebí operace join, která se při použití ve vyšším množství může stát nevýhodou.

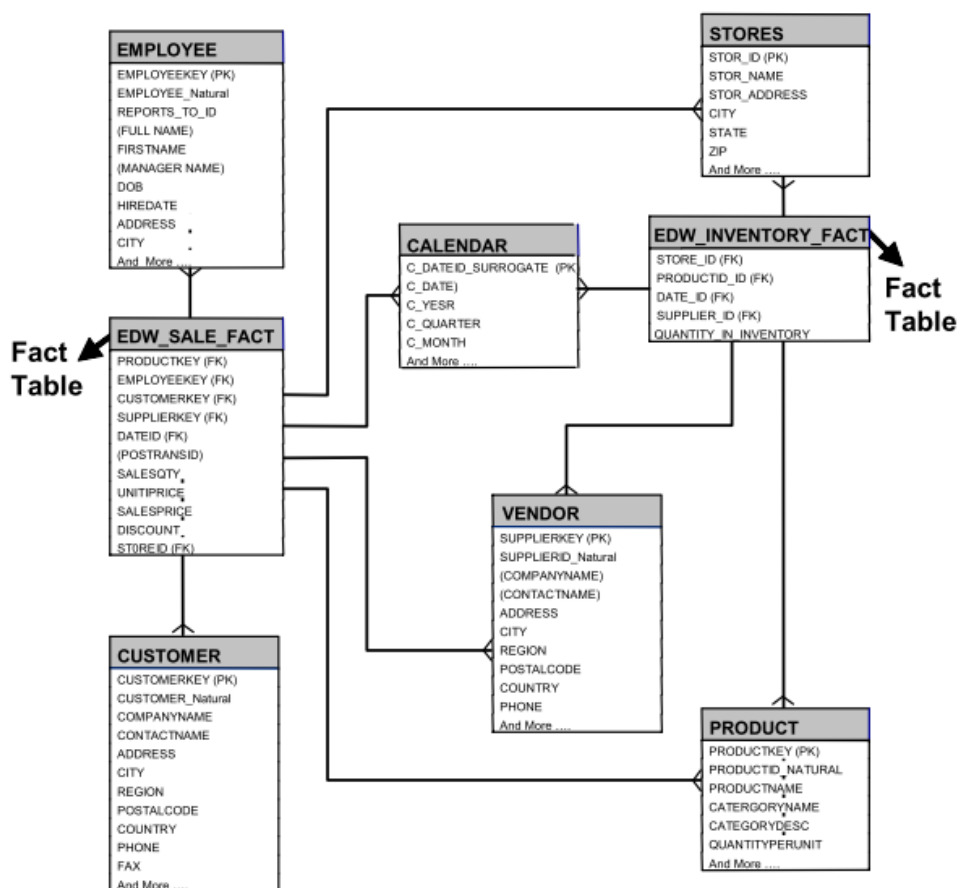


Obr. 2. 9 - Schéma snowflake (sněhové vločky)

[Zdroj: vlastní]

Datový model však může mít podobu i dalších multidimenzionálních schémat. Jedním z nich je např. multi-star schéma.

Multi-star schéma je dimenzionální schéma obsahující několik tabulek faktů spojených mezi sebou pomocí tabulek dimenzí.



Obr. 2. 10 - Multi-star schéma

[Zdroj: <http://www.redbooks.ibm.com>]

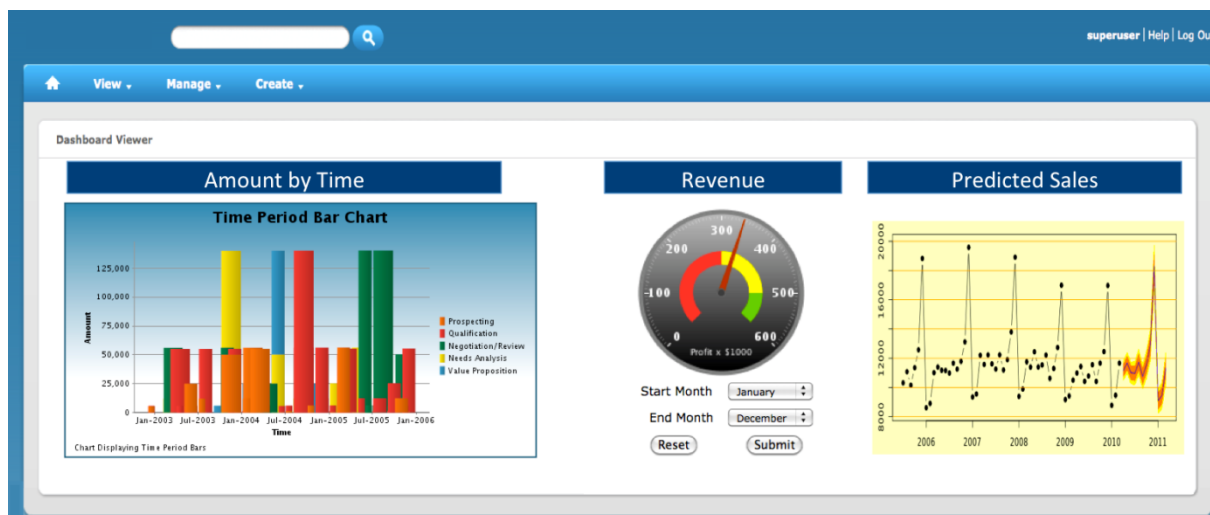
2.6. Dashboard - vizuální zobrazení BI řešení

Dashboard, česky také palubní či přístrojová deska, je nástrojem, který poskytuje vrcholovým pracovníkům společnosti účinnou podporu při jejich každodenní práci a při operativním i strategickém rozhodování.

Few (2006) definuje dashboard jako vizuální zobrazení nejdůležitějších informací potřebných k dosažení jednoho nebo více cílů společnosti, které je konsolidované a uspořádané na jediné obrazovce, což umožňuje sledovat klíčové informace na první pohled.

2.6.1. Základní charakteristiky správně navrženého dashboardu

- Vizuální zobrazení: informace na dashboardu jsou prezentovány vizuálně, obvykle jako kombinace textu a grafů. Důraz je kladen především na grafy, jelikož grafická prezentace informací je jasnější a snáze pochopitelná než text samotný.
- Konkrétní cíl: dashboard zobrazuje často kolekci informací z různých zdrojů, které společně vedou k dosažení konkrétního cíle (společnosti).
- Jediná obrazovka počítače: informace na dashboardu by měly být na jediné stránce - obrazovce tak, aby byl celý obsah viditelný najednou na první pohled. To vede k tomu, že nejdůležitější informace, které jsou potřeba, jsou rychle a snadno k dispozici.
- Konsolidace informací: navzdory tomu, že by dashboard měl zobrazovat klíčové informace na jediné obrazovce, je nezbytné, aby byly zpřístupněny i detailnější informace, upozorňující na problémy či nedostatky v chodu společnosti.
- Stručné, jasné a intuitivní mechanismy prezentující informace: na dashboardu by se měly objevovat pouze takové prezentační mechanismy, které jasně vystihují obsah sdělované informace.
- Připraven „na míru“: dashboardy jsou kustomizovány tak, aby informace, které obsahují, byly přesně přizpůsobeny požadavkům dané osoby, skupině osob nebo společnosti.



Obr. 2. 11 - ukázka dashboardu

[Zdroj: <http://www.jaspersoft.com>]

2.6.2. Efektivní zobrazovací média dashboardu

Dashboard musí být založen na vhodně vybraných zobrazovacích médiích tak, aby bylo dosaženo jeho potenciálu s cílem jasného a bezprostředního zobrazení konsolidovaných informací o společnosti.

Mezi základní zobrazovací média patří:

- grafy,
- ikony,
- tabulky,
- mapy.

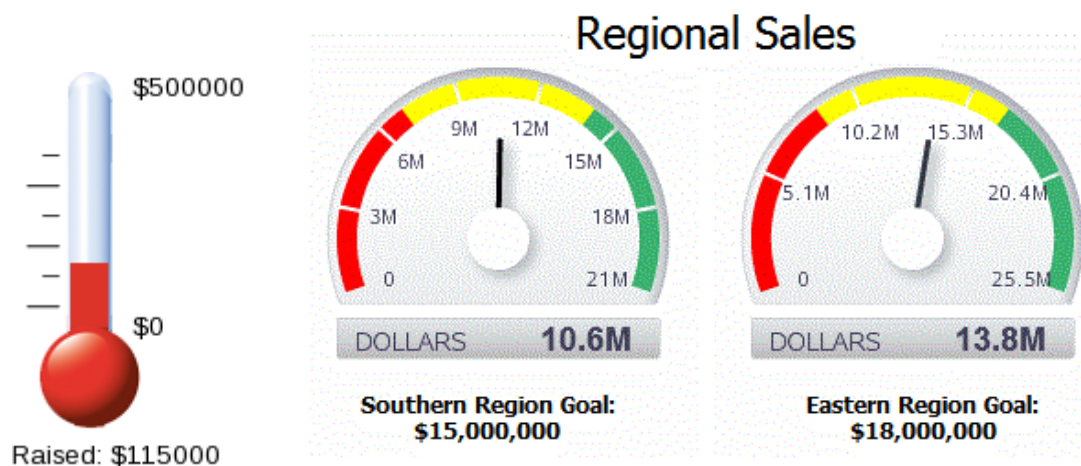
a. Grafy

Vzhledem k tomu, že obsahem většiny dashboardů jsou kvantitativní data, jsou grafy jedním z nejčastěji používaných zobrazovacích médií.

Použitelné jsou téměř veškeré typy grafů, jako je např.:

- sloupcový graf (horizontální, vertikální, ale také varianta skládaného sloupcového grafu),
- spojnicový graf,
- kombinace sloupcového a spojnicového grafu,
- graf pro korelační analýzy.

Speciálním typem zobrazovacího média, které bychom mohli také zařadit do kategorie grafů, jsou různé typy měřidel. Takový graf může mít podobu např. teploměru, či tachometru, který zobrazuje nejčastěji pouze jeden ukazatel. Někdy je tento ukazatel zobrazen v porovnání s hodnotou, kterou chce společnost dosáhnout. Tak je na první pohled viditelné, jestli si vede dobře, nebo špatně.



Obr. 2. 12 - ukázka speciálního typu grafů

[Zdroj: <http://docs.oracle.com>]

b. Ikony

Často je důležité, aby bylo včas upozorněno na konkrétní informace na dashboardu, zejména pokud je něco v nepořádku a vyžaduje to pozornost uživatele.

V těchto případech se často využívá ikon v podobě semaforů tak, jak je známe v dopravě. Zelená barva obvykle signalizuje, že je vše v pořádku. Naopak oranžová, resp. častěji žlutá barva a červená barva upozorňují na možný problém.

Dalším příkladem ikon jsou tzv. up/down ikony, které jednoduchým způsobem vizualizují, zda se sledovaný ukazatel zvyšuje či snižuje v porovnání např. s minulým měsícem, rokem apod.

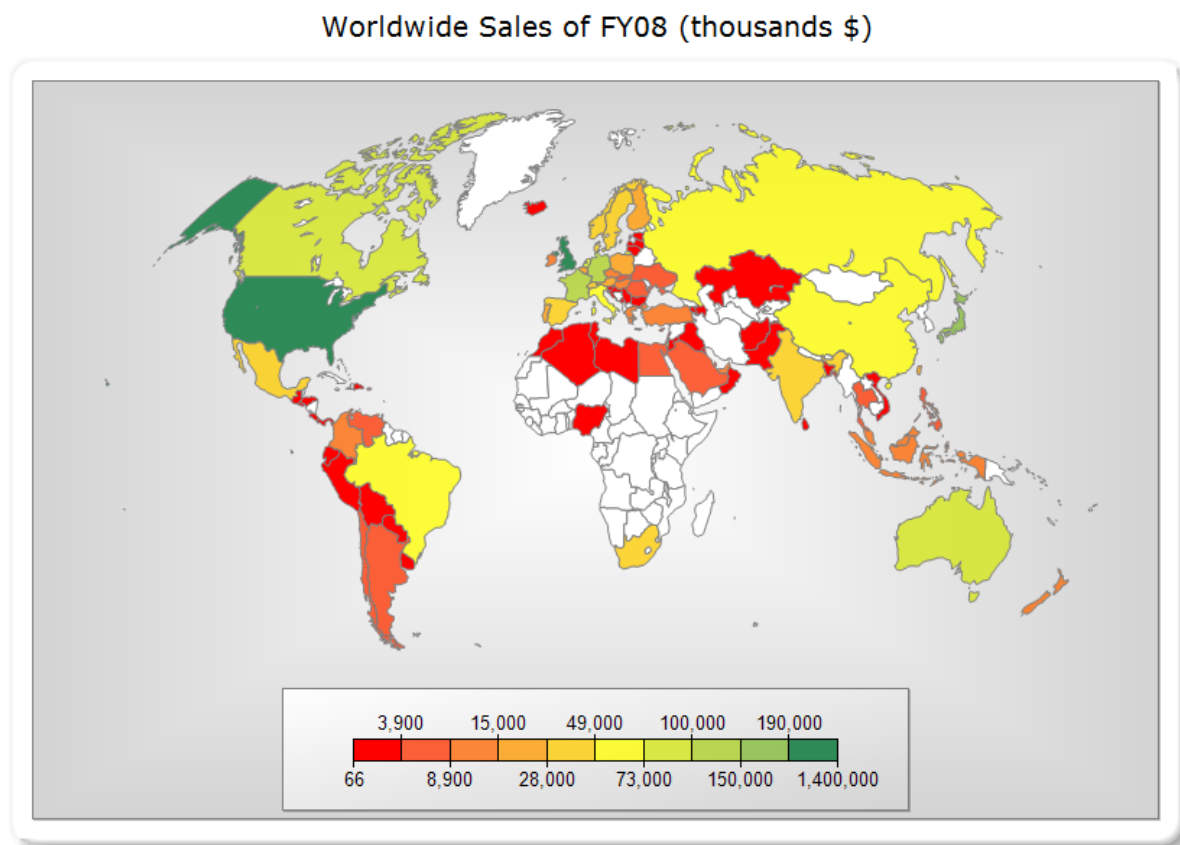
c. Tabulky

Tabulky jsou dalším velice často využívaným a oblíbeným médiem pro vizualizaci dat na dashboardu.

Jejich výhodou je přehlednost a přesné zobrazení hodnot jednotlivých ukazatelů, avšak při velkém množství sloupců či řádků se může stát nepřehlednou a uživateli tak nepoužívanou.

d. Mapy

Mapy nabízí více specializovanou a méně často potřebnou formu vizualizace. Jsou obvykle používány k přiřazení jednotlivých dat ke konkrétní části prostoru na mapě. Tím je v mnoha případech docíleno zvýšení významu těchto dat a je tak možno rychle reagovat na vývoj ukazatele v jednotlivých částech na mapě.



Obr. 2. 13 - ukázka mapy jako zobrazovací médium BI

[Zdroj: <http://blogs.technet.com>]

Dashboards umožňují nejen monitorovat aktuální výkonnost jednotlivých obchodních procesů společnosti, identifikovat kritické ukazatele a zkoumat je v požadované úrovni detailu, ale také pomocí analýz typu „Co se stane, když...“ modelovat situace bez nutnosti programování a hledat možné příležitosti ke zlepšení výkonů nebo např. vhodnou úsporu nákladů.

2.7. Shrnutí

V teoretické části práce byly čtenáři přiblíženy jednotlivé pojmy z oblasti Business Intelligence a představeny vrstvy a komponenty. Dále zde byl popsán princip ukládání dat, neboli multidimenzionalita, a dimenzionální modelování v řešeních BI. Na závěr teoretické části byla nastíněna problematika z oblasti vizualizace řešení Business Intelligence.

Na první část práce, s využitím teorie v ní obsažené, bude navazovat aplikační část, kde bude navrženo řešení pro prezentaci manažerských informací formou dashboardu. Současně bude také navržen způsob čerpání dat ze zdrojových systémů (ETL proces) a způsob jejich uložení.

3. Analýza výchozí stavu řešení prezentace prodejních výsledků

3.1. Představení společnosti

Společnost, pro kterou je navrhován dashboard pro prezentaci prodejních výsledků, působí na mezinárodním trhu v oblasti výroby papíru pro tisk novin, knih a časopisů. Dále také nabízí papír vysoké kvality, lepenku, průmyslové obaly a široký sortiment výrobků ze dřeva. Jedná se o akciovou společnost, která je veřejně obchodována především na burzách v Helsinkách a Stockholmu a zaměstnává více než 27 tisíc pracovníků v 88 výrobních závodech, rozmístěných v 35 zemích světa.

3.2. Výchozí stav řešení prezentace prodejních výsledků

V současné době je prezentace prodejních výsledků řešena jako export dat za určité období, pro určité strategické podnikatelské jednotky společnosti, ze zdrojových systémů do nástroje Microsoft Excel. Zde je prováděna analýza dat a tvorba grafů prezentující prodejní výsledky dle požadavků managementu společnosti.

Hlavním nedostatkem je nutnost manuálně každý měsíc provést export dat ze zdrojových systémů a poté vytvořit nové analýzy, popřípadě aktualizovat ty z minulého měsíce. Než se výsledky analýz dostanou k managementu společnosti, trvá to obvykle déle, než je žádoucí. Důvodem je nutnost zapojit do tohoto procesu mnoho uživatelů. Pro ilustraci: jedna skupina uživatelů exportuje data z jednotlivých zdrojových systémů, další skupina provádí jednotlivé analýzy, které management očekává na začátku každého měsíce. Dalším problémem je distribuce těchto analýz mezi členy managementu. V současné době se tak děje prostřednictvím emailů, což není zcela spolehlivá a efektivní metoda doručení, jak mi bylo potvrzeno členy managementu.

Nedostatky budou odstraněny po nasazení navrhovaného dashboardu. V řešení bude využíván ETL proces, který eliminuje potřebu manuálně cokoli exportovat a dále bude využívána OLAP databáze, kde se budou automaticky provádět veškeré analýzy. Management tak bude mít veškeré výsledky analýz přehledně uspořádané na jediném dashboardu, ke kterému se dostanou prostřednictvím intranetových stránek společnosti a odpadne tak nutnost používání emailů k doručení.

Řešení navrhovaného dashboardu se bude opírat o současnou podobu grafů, které reprezentují výsledky analýz. Jejich struktura a koncepce se jeví jako velice optimální a tedy, podle slov managementu, efektivní pro práci s nimi.

Ke grafům dále přibudou nové analýzy, které v současné podobě řešení nebylo možné prezentovat, jelikož nástroj Microsoft Excel neobsahoval veškeré prostředky, které jsou k tomu nutné.

4. Návrh řešení dashboardu pro prezentaci prodejních výsledků

4.1. Sběr požadavků

Sběr požadavků, na návrh dashboardu pro prezentaci prodejních výsledků společnosti, byl proveden metodou skupinového rozhovoru. Jednotliví respondenti byli zástupci z různých pozic managementu společnosti.

Cílem tohoto rozhovoru bylo získat celkový přehled o společnosti, její vize a cíle, způsob měření úspěšnosti a její postavení na trhu. Dále bylo nutné identifikovat klíčové ukazatele, jež management využíval k řízení společnosti, resp. dané strategické podnikatelské jednotky.

Výsledkem skupinového rozhovoru je tabulka 4.1 se všemi uživatelskými požadavky na výsledný dashboard pro prezentaci prodejních výsledků společnosti.

Uživatelské požadavky			
číslo	Charakteristika	Období	Datový zdroj
1.	Prezentace produkce v jednotlivých strategických podnikatelských jednotkách	vývoj v minulém roce+letošní rok+odhad na příštích 12 měsíců	zdrojový systém/export dat z plánovacího (reportovacího) systému
2.	Prezentace aktuální, tzn. denní situace produkce za celou podnikatelskou oblast (pro kterou je navrhován dashboard)	každodenní aktualizace	zdrojový systém
3.	Prezentace stavu objednávek za celou podnikatelskou oblast	vývoj v minulém roce+letošní rok	zdrojový systém
4.	Prezentace vyfakturovaného objemu produktů za jednotlivé podnikatelské jednotky	vývoj v minulém roce+letošní rok+odhad na příštích 12 měsíců	zdrojový systém/export dat z plánovacího (reportovacího) systému
5.	Prezentace vyfakturovaného objemu produktů v jednotlivých měnách	vývoj v minulém roce+letošní rok	zdrojový systém
6.	Prezentace změny prodeje v zemích, kam se exportují vyrobené produkty	srovnání minulého a letošního roku	zdrojový systém

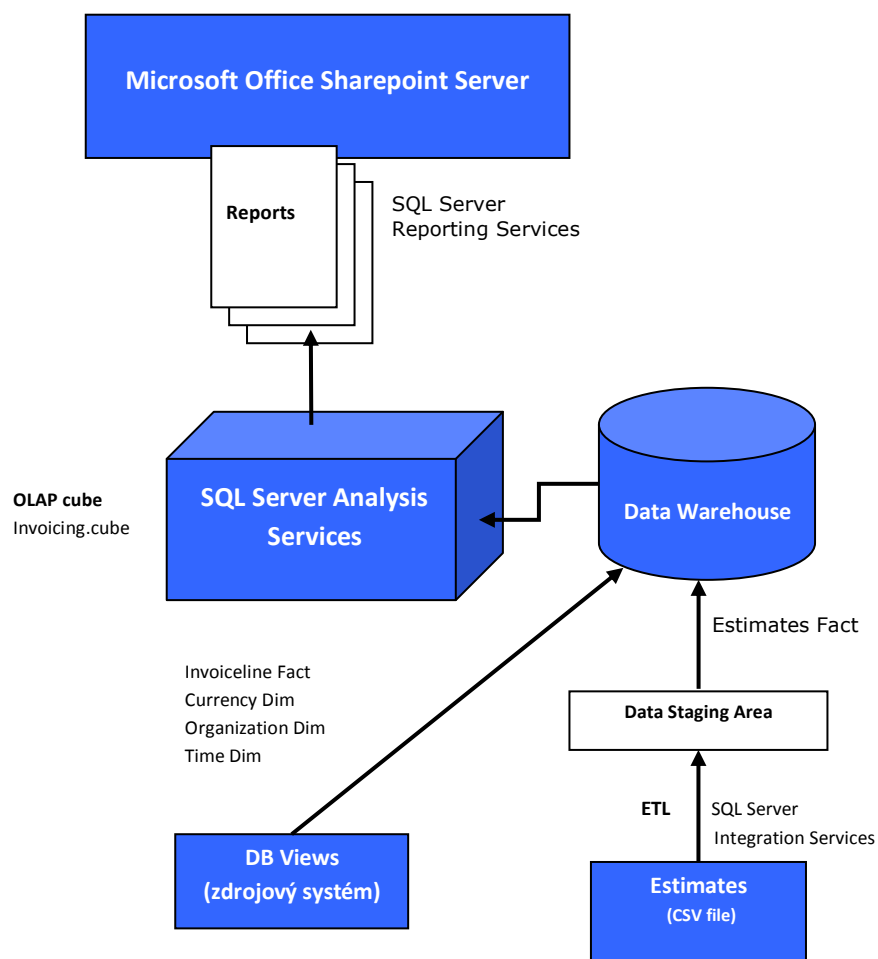
Tab. 4. 1 - Uživatelské požadavky

4.2. Analýza uživatelských požadavků

Analýzou uživatelských požadavků bylo zjištěno, že klíčové ukazatele úspěšnosti prodejních výsledků jsou: výše vyrobených produktů, resp. produkce jednotlivých strategických podnikatelských jednotek, a výše vyfakturovaného objemu produktů. Dalšími ukazateli, které vyplynuly jako důležité pro management společnosti, jsou změny prodeje v zemích, kam se exportují produkty dané společnosti, a také stav jednotlivých objednávek.

Architektura řešení návrhu dashboardu pro prezentaci prodejních výsledků společnosti bude následující:

- faktová data zachycující odhad na příští období (Estimates) budou procesem ETL extrahována do dočasného uložení dat (Data Staging Area), kde budou transformována a nahrána do datového skladu (Data Warehouse),
- faktová data za minulé období až do současnosti budou v datovém skladu (Data Warehouse) uložena jako snímky ze zdrojového systému,
- na základě datového skladu bude poté vytvořena OLAP databáze (OLAP kostka) s dvěma tabulkami faktů a třemi tabulkami dimenzí,
- reporty budou vytvořeny s využitím dat z OLAP kostky,
- výsledný dashboard bude umístěn na Microsoft Office Sharepoint Serveru.



Obr. 4. 1 - Architektura řešení pro návrh dashboardu

[Zdroj: vlastní]

4.3. Návrh, modelování a implementace řešení

Celé řešení pro návrh dashboardu k prezentaci prodejních výsledků bude vytvořeno s použitím nástrojů firmy Microsoft. Konkrétně se jedná o tyto nástroje:

- SQL Server Management Studio,
- SQL Server Business Intelligence Development Studio:
 - SQL Server Integration Services,
 - SQL Server Analysis Services,
 - SQL Server Reporting Services,
- Microsoft Office Sharepoint Server.

4.3.1. Dimenzionální modelování

Z analýzy uživatelských požadavků vyplynulo, že v datovém skladu, resp. v OLAP kostce budou využívány tyto dimenze:

- Dimenze pro organizační strukturu

Dimenze: ORGANIZATION	
Hierarchie	BusinessArea BusinessUnit MillBusinessUnit ProductSegment Product

V dimenzi ORGANIZATION, zachycující organizační strukturu společnosti, bude vytvořena hierarchie dle výše uvedené tabulky, kde:

- Business Area = podnikatelská oblast,
- Business Unit = (strategická) podnikatelská jednotka,
- Mill Business Unit = podnikatelská subjednotka,
- Product segment = část produktů vyráběných v subjednotkách,
- Product = konkrétní produkt.

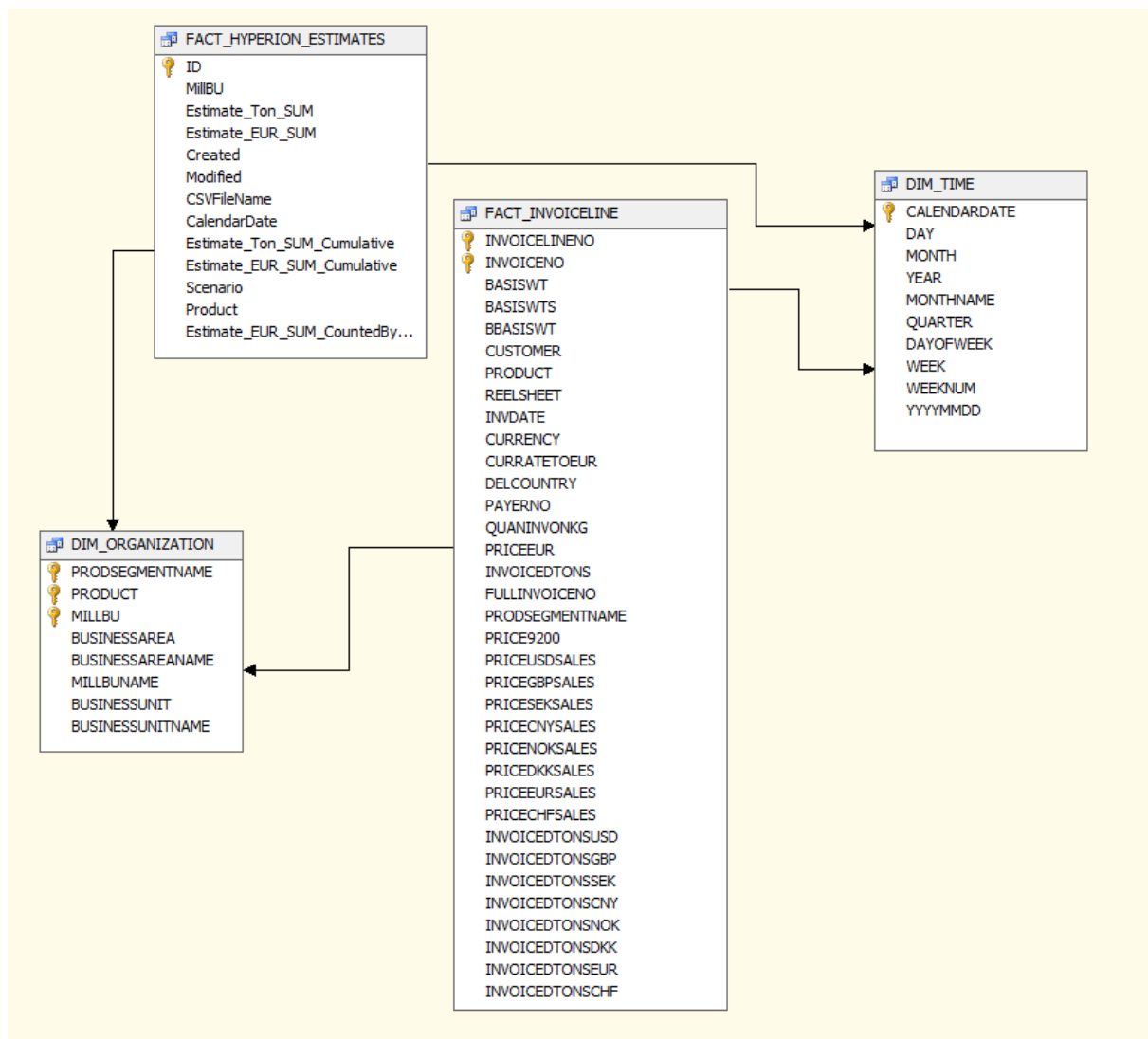
- Časová dimenze

Dimenze: TIME	
Hierarchie	Year Quarter Month Week Day

Časová dimenze je rozdělena do obvyklé hierarchie a to: Rok – Čtvrtletí – Měsíc – Týden - Den.

4.3.2. Model datového skladu

Z dimenzionálního modelování vyplynul následující model datového skladu, který je zachycen na níže uvedeném obrázku.



Obr. 4. 2 - datový model datového skladu

[Zdroj: vlastní]

Datový sklad bude obsahovat dvě tabulky faktů. První bude obsahovat odhad (plán) na následující období, který bude procesem ETL nahráván z plánovacího (reportovacího) systému společnosti. Druhá faktová tabulka bude obsahovat jak historické, tak také aktuální hodnoty - fakta.

Další část datového skladu tvoří tabulky dimenzí. Jedná se o časovou dimenzi a dimenzi zachycující organizační strukturu společnosti.

4.3.3. Návrh a implementace ETL procesu

Do ETL procesu je nutné zakomponovat také návrh dočasného uložení dat (Data Staging Area), jelikož extrahované data z plánovacího (reportovacího) systému Oracle Hyperion je nutné před nahráním do datového skladu ještě upravit.

a. Data Staging Area (Dočasné uložení dat)

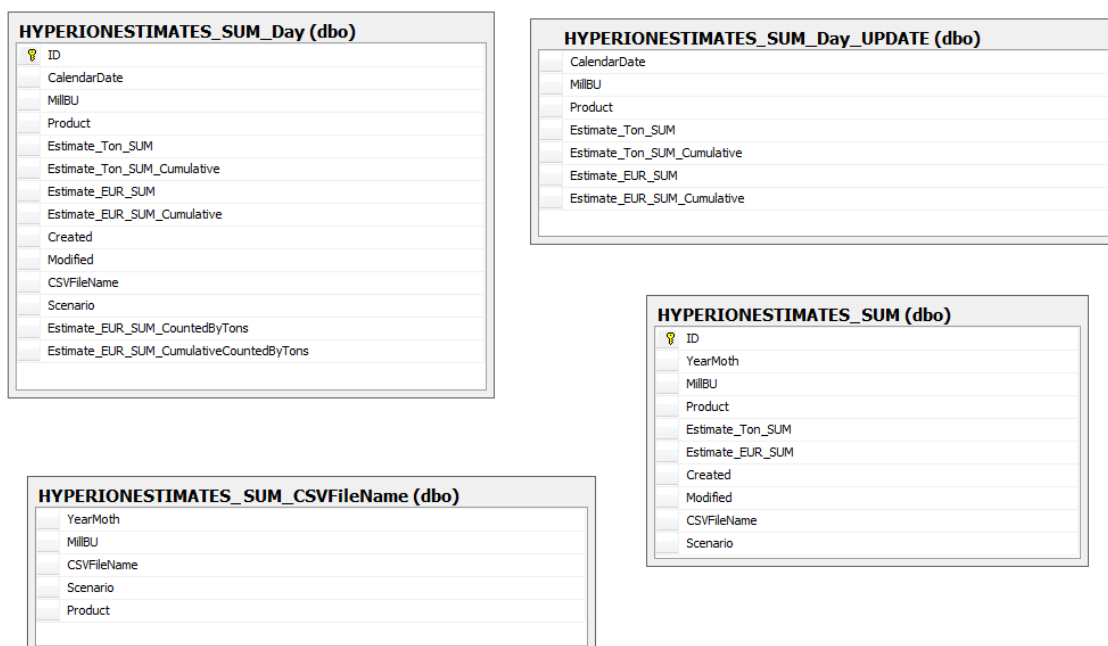
Dočasné uložení dat je tvořeno tabulkami, do kterých jsou uloženy upravené data z plánovacího systému. Tabulky budou mít následující strukturu:

HYPERIONESTIMATES_SUM_Day			
Atribut	Typ	Nulový	Primární klíč
ID	int	ne	ano
CalendarDate	datetime	ne	ne
MillBU	varchar (4)	ne	ne
Product	varchar (50)	ano	ne
Estimate_Ton_SUM	decimal (18, 2)	ano	ne
Estimate_Ton_SUM_Cumulative	decimal (18, 2)	ano	ne
Estimate_EUR_SUM	decimal (18, 2)	ano	ne
Estimate_EUR_SUM_Cumulative	decimal (18, 2)	ano	ne
Created	datetime	ano	ne
Modified	datetime	ano	ne
CSVFileName	varchar (200)	ano	ne
Scenario	varchar (50)	ano	ne
Estimate_EUR_SUM_CountedByTons	decimal (18, 2)	ano	ne
Estimate_EUR_SUM_CumulativeCountedByTons	decimal (18, 2)	ano	ne

HYPERIONESTIMATES_SUM			
Atribut	Typ	Nulový	Primární klíč
ID	int	ne	ano
YearMonth	int	ne	ne
MillBU	varchar (4)	ne	ne
Product	varchar (50)	ano	ne
Estimate_Ton_SUM	decimal (18, 2)	ano	ne
Estimate_EUR_SUM	decimal (18, 2)	ano	ne
Created	datetime	ano	ne
Modified	datetime	ano	ne
CSVFileName	varchar (200)	ano	ne
Scenario	varchar (50)	ano	ne

HYPERIONESTIMATES_SUM_Day_UPDATE			
Atribut	Typ	Nulový	Primární klíč
CalendarDate	datetime	ne	ne
MillBU	varchar (4)	ne	ne
Product	varchar (50)	ano	ne
Estimate_Ton_SUM	decimal (18, 2)	ano	ne
Estimate_Ton_SUM_Cumulative	decimal (18, 2)	ano	ne
Estimate_EUR_SUM	decimal (18, 2)	ano	ne
Estimate_EUR_SUM_Cumulative	decimal (18, 2)	ano	ne

HYPERIONESTIMATES_SUM_CSVFileName			
Atribut	Typ	Nulový	Primární klíč
YearMonth	int	ano	ne
MillBU	varchar (4)	ano	ne
CSVFileName	varchar (200)	ano	ne
Scenario	varchar (50)	ano	ne
Product	varchar (50)	ano	ne



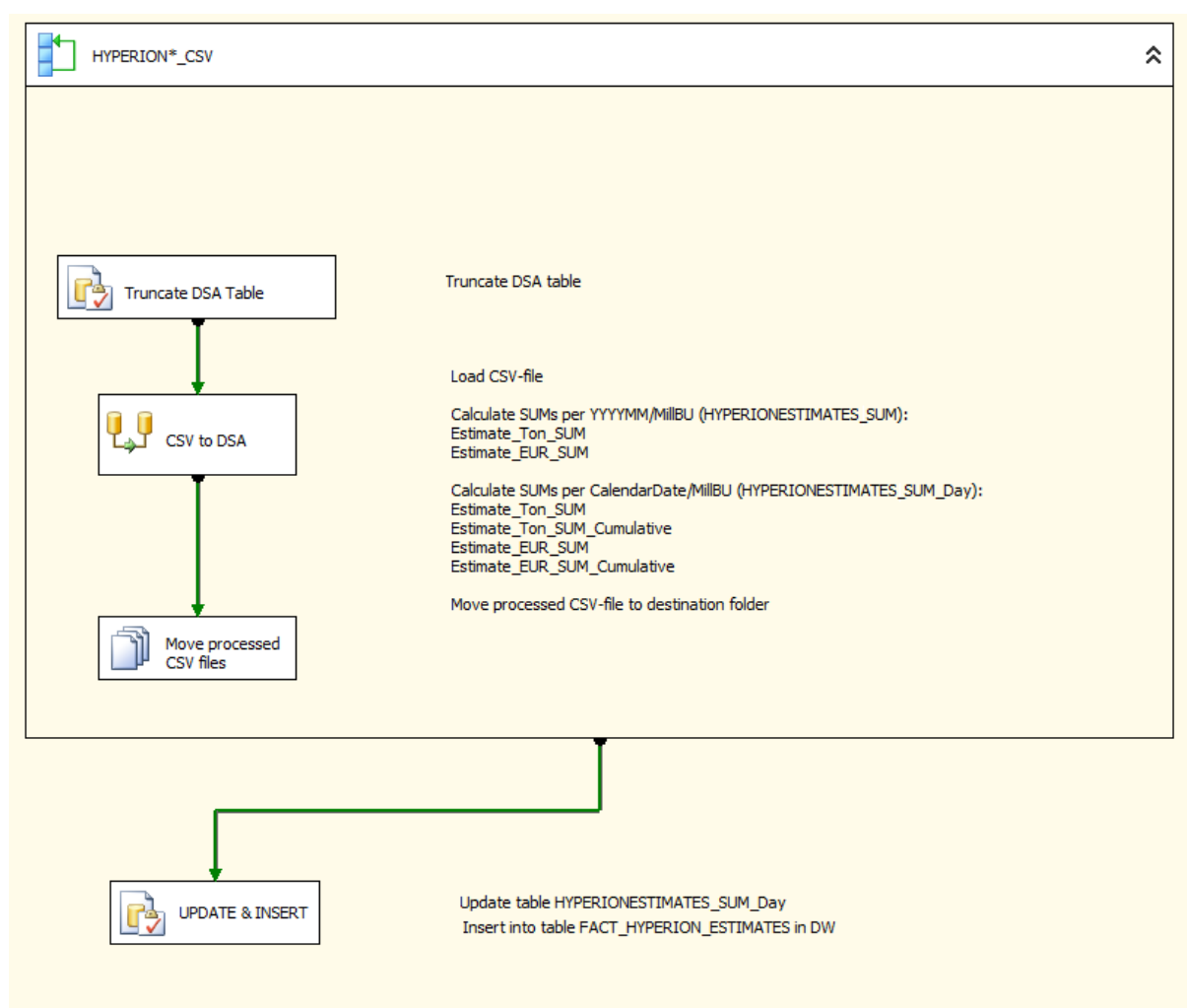
Obr. 4. 3 - model dočasného uložště dat

[Zdroj: vlastní]

b. Datová pumpa

Cílem datové pumpy je nahrát data z exportovaného CSV (comma-separated values) souboru z plánovacího systému do dočasného uložení dat, provést transformaci a poté upravená data nahrát do datového skladu.

Na začátku ETL procesu jsou vymazány záznamy v tabulkách dočasného uložení dat (DSA). Poté se řádek po řádku CSV souboru nahraje do příslušných tabulek DSA. Jakmile jsou načteny všechny řádky, začne samotná transformace dat.



Obr. 4. 4 - ETL proces

[Zdroj: vlastní]

První proces probíhající v rámci transformace dat je změna desetinné tečky za desetinnou čárku a převod těchto hodnot na datový typ *Decimal*. Dále se ETL proces rozdělí do dvou větví.

První větev vloží do DSA tabulky *HYPERIONESTIMAS_SUM_CsvFileName* název CSV souboru, aby v budoucnu bylo možné identifikovat naposledy zpracovaný soubor.

Druhá větev rozdělí data podle příznaku ze sloupce *Account*, zda se jedná o peněžní hodnoty (cena), nebo o množstevní hodnoty (tuny) a na základě toho je přejmenován sloupec, do kterého se tyto hodnoty ukládají, aby bylo možné identifikovat, o jakou hodnotu se jedná. Následně jsou obě tyto skupiny dat datumově seřazeny a opět sloučeny do jediné skupiny.

V další části ETL procesu je nutné rozdělit a vložit data do příslušných DSA tabulek a to na základě toho, zda se jedná o data sumarizované za měsíc, nebo denní data.

Pokud se jedná o denní data je proces následující: nejprve se zkontroluje, zda je chronologicky seřazena tabulka kalendáře. Poté se k denním hodnotám přidá konkrétní datum na základě propojení klíčů *CalendarDate* a provede se kalkulace kumulativních hodnot *Estimate_Ton_SUM_Cumulative* a *Estimate_EUR_SUM_Cumulative*. Posledním krokem této větve ETL procesu je vyhledat a porovnat jednotlivé záznamy tabulky *HYPERIONESTIMATES_SUM_Day* a pokud se zde vkládaný záznam nenachází, vložit jej. Pokud je vkládaný záznam v tabulce nalezen, ale má jinou hodnotu, je vložen do tabulky *HYPERIONESTIMATES_SUM_Day_UPDATE*. Z této tabulky je poté zkopírován až jako poslední krok z důvodu zlepšení výkonu a rychlosti datové pumpy.

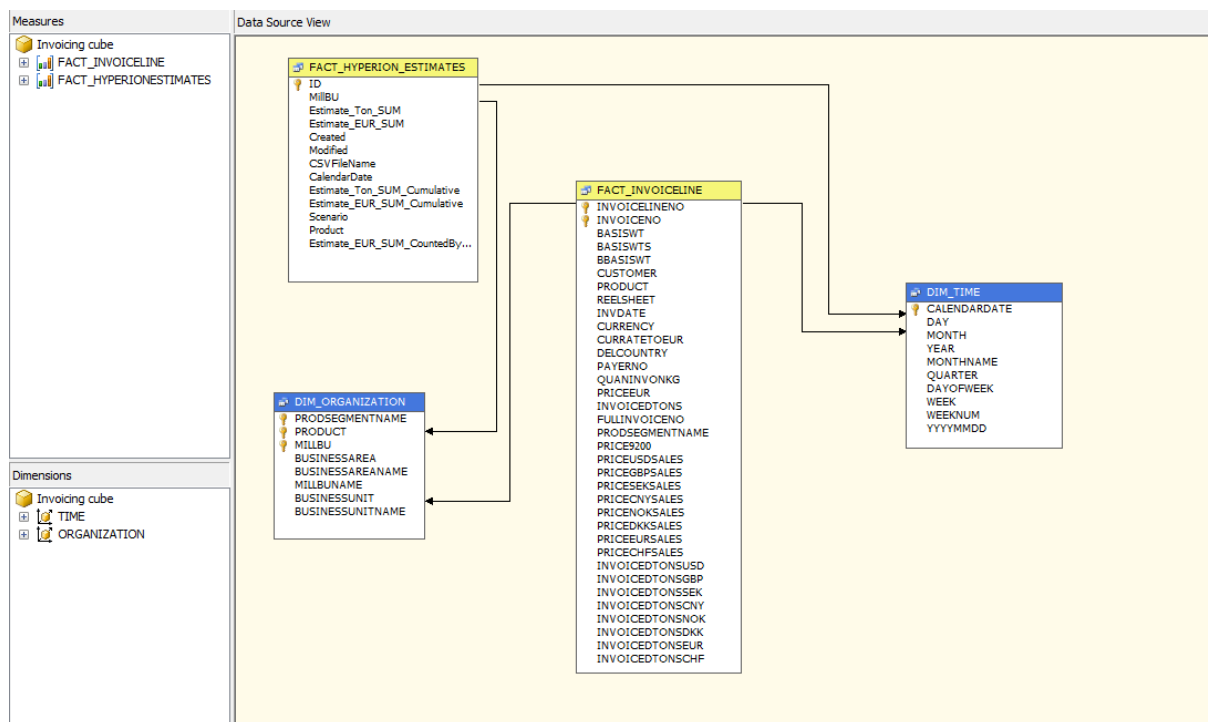
Druhá větev zpracovávající měsíční hodnoty porovná, zda vkládaný záznam se již v tabulce *HYPERIONESTIMATES_Sum* nachází, či nikoliv. Na základě toho je buď konkrétní záznam vložen, pak se jedná o nový záznam, nebo je v tabulce nalezen a pak je aktualizována jeho hodnota a přidáno datum o jeho modifikaci.

Posledním krokem datové pumpy je aktualizace denních dat, které, když bylo prováděno v rámci procesu transformace dat, značně zpomalovalo celý ETL proces. V tomto kroku se aktualizují data v tabulce *HYPERIONESTIMATES_SUM_Day* na základě dat z tabulky *HYPERIONESTIMATES_SUM_Day_UPDATE* a přidá se datum o provedené modifikaci. Nakonec se záznamy z tabulek dočasného uložení dat vloží do datového skladu.

Obrázek ilustrující proces transformace dat je umístěn mezi přílohami na konci této práce.

4.3.4. Návrh a implementace OLAP databáze

OLAP databáze je vytvořena na základě modelu datového skladu, to znamená, že OLAP kostka, stejně jako datový sklad, bude obsahovat dvě tabulky faktů a dvě tabulky dimenzí, které jsou popsány v části zabývající se modelem datového skladu.



Obr. 4. 5 - model OLAP databáze

[Zdroj: vlastní]

4.3.5. Návrh reportů pro prezentaci řešení

a. *Produkce v jednotlivých podnikatelských jednotkách – trend*

Navržený report zachycuje trend, jaký měla produkce v jednotlivých podnikatelských jednotkách v minulém roce, letošním roce a zároveň je zde zachycena plánovaná produkce na následujících 12 měsíců. Report je aktualizován jednou za týden, plánované hodnoty na příštích 12 měsíců jsou nahrávány na začátku každého nového měsíce.

Shrnutí reportu:

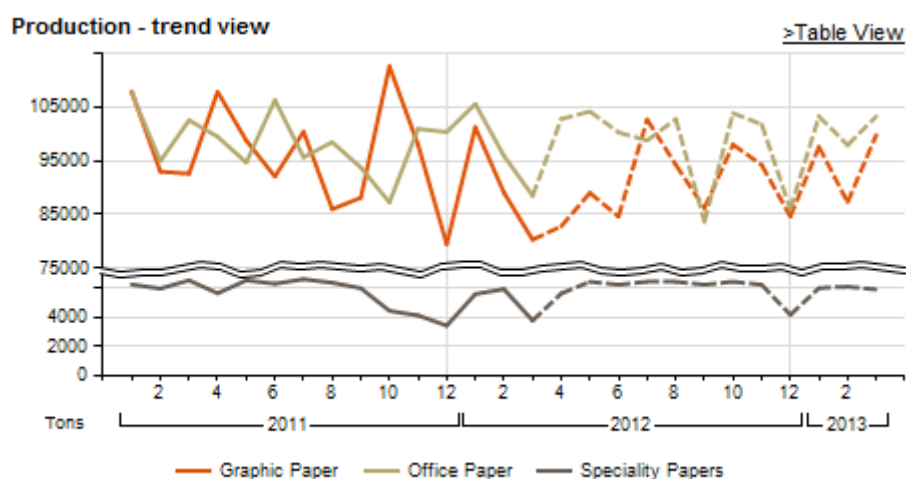
Plná čára: skutečná produkce

Prerušovaná čára: odhad podle posledního plánu na příštích 12 měsíců

Osa X: měsíce, roky

Osa Y: tuny

Aktualizace reportu: jednou týdně, plán na další období je nahráván jednou měsíčně



Obr. 4. 6 - Report - Produkce v jednotlivých podnikatelských jednotkách (trend)

[Zdroj: vlastní]

b. Denní produkce za celou podnikatelskou oblast

Tento report zachycuje opět produkci, ale na rozdíl od předešlého reportu, souhrnně za všechny podnikatelské jednotky pro daný den, tedy denní stav produkce za celou podnikatelskou oblast.

Shrnutí reportu:

Skutečný denní stav produkce

Osa: tuny x 1000

Aktualizace reportu: jednou denně

Production - today view



Obr. 4. 7 - Report - Denní produkce za celou podnikatelskou oblast

[Zdroj: vlastní]

c. Stav objednaného zboží

Stav objednaného zboží je zachycen v reportu, který se skládá ze sloupcového grafu a spojnicového grafu.

Sloupcový graf je složen z modré a zelené části. Modrá část zachycuje objem produkce, který je vyráběn v rámci objednávek, na které již čekají zákazníci. Zelená část zobrazuje objem produkce, který je vyráběn už nad rámec objednávek, tedy do zásoby.

Trend toho, zda se vyráběly produkty pouze na objednávky či do zásoby v jednotlivých dnech, je zachycen spojnicovým grafem.

Shrnutí reportu:

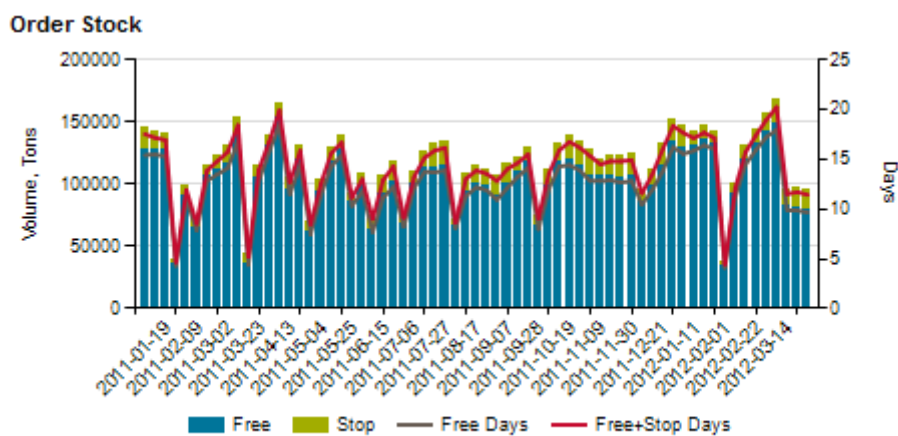
Sloupcový graf: vyrobený objem produktů v tunách

Spojnicový graf: počet dní, kdy jsou stroje schopny vyrábět na objednávky/do zásoby

Osa X: datum

Osa Y: objem produkce v tunách; dny

Aktualizace reportu: jednou denně



Obr. 4. 8 - Report - Stav objednaného zboží

[Zdroj: vlastní]

d. Vyfakturovaný objem produktů za jednotlivé podnikatelské jednotky

Report zachycuje trend, jaký měl vyfakturovaný objem produktů v jednotlivých podnikatelských jednotkách v minulém roce, v letošním roce a zároveň je zde zobrazen plánovaný objem vyfakturovaných produktů na následujících 12 měsících. Report je aktualizován denně, plánované hodnoty na příštích 12 měsících jsou nahrávány na začátku každého měsíce.

Shrnutí reportu:

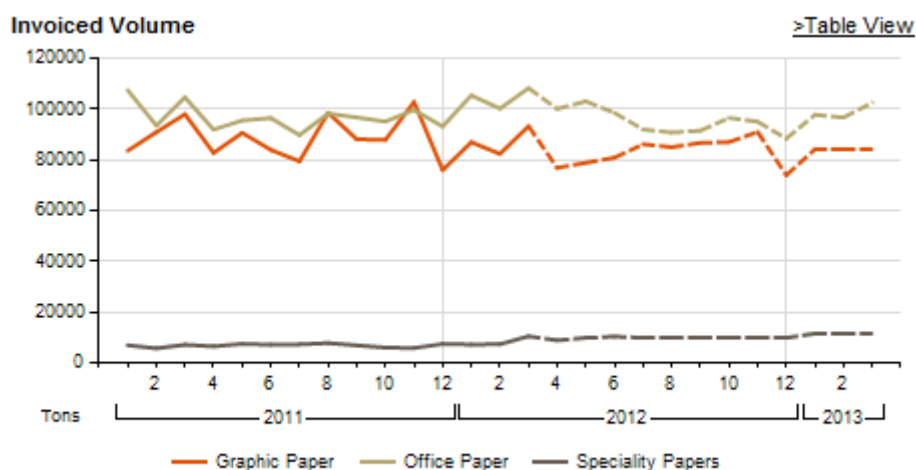
Plná čára: skutečné množství vyfakturovaného objemu produkce

Přerušovaná čára: odhad podle posledního plánu na příštích 12 měsících

Osa X: měsíce, roky

Osa Y: tuny

Aktualizace reportu: denně, plánované hodnoty jsou nahrávány jednou měsíčně



Obr. 4. 9 - Report - Vyfakturovaný objem produktů v jednotlivých podnikatelských jednotkách

[Zdroj: vlastní]

e. Vyfakturovaný objem produktů za jednotlivé světové měny

Report zachycuje objem produktů v jednotlivých podnikatelských jednotkách, rozdělený po měnách v jakých byl objem produktů vyfakturován, v minulém roce a v letošním roce pomocí sloupčového grafu. Dále je zde zachycen přerušovanou čarou celkový vyfakturovaný objem produktů v eurech. Tento report je aktualizován denně.

Shrnutí reportu:

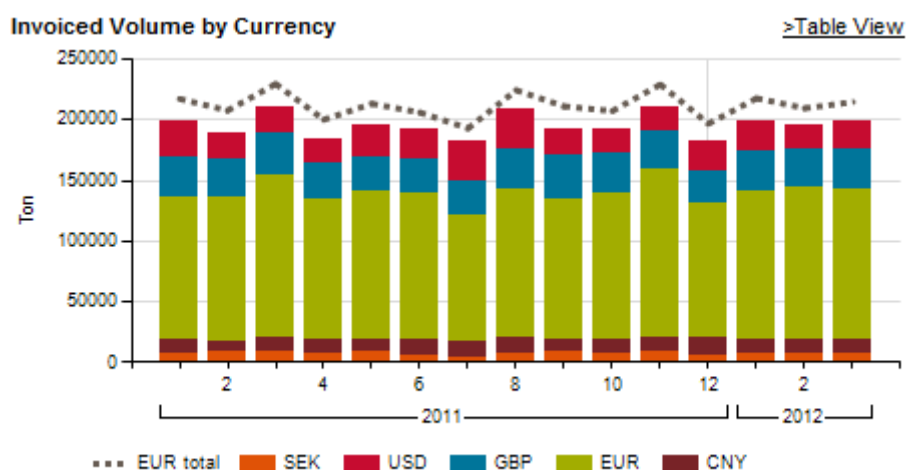
Sloupčový graf: skutečné množství vyfakturovaného objemu produkce podle měn

Přerušovaná čára: celkové množství vyfakturovaného objemu produkce v eurech

Osa X: měsíce, roky

Osa Y: tuny

Aktualizace reportu: denně



Obr. 4. 10 - Report - Vyfakturovaný objem produktů za jednotlivé světové měny

[Zdroj: vlastní]

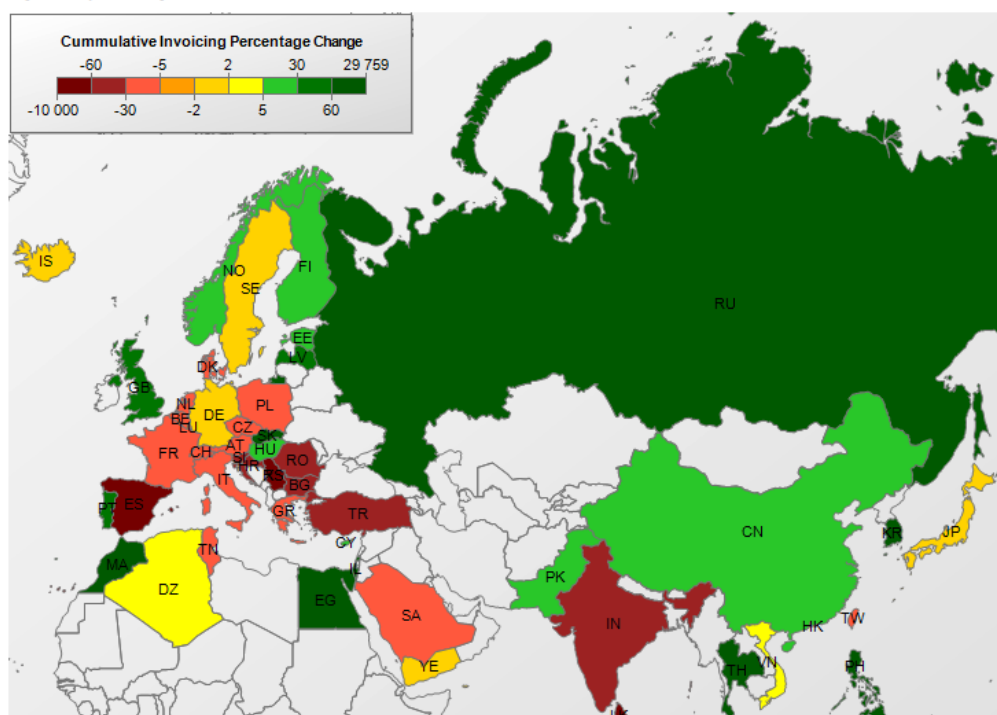
f. Změna prodeje v zemích, kam se exportují vyrobené produkty

Posledním reportem je report zobrazující změnu prodeje v zemích, kam se exportují vyrobené produkty společnosti. Je zde porovnáván objem prodeje v letošním a loňském roce. Zelená barva na stupnici znázorňuje, že se objem prodeje oproti minulému roku zvýšil. Naopak červená barva znamená, že se objem prodeje v letošním roce v dané zemi snížil.

Shrnutí reportu:

Aktualizace reportu: čtvrtletně

Sales change map - current year vs. previous year



Obr. 4. 11 - Report - Změna prodeje v zemích, kam se exportují vyrobené produkty

[Zdroj: vlastní]

4.4. Nasazení řešení do provozu

Na předchozí fázi, modelování a návrh dashboardu pro prezentaci prodejních výsledku společnosti, bude navazovat poslední fáze a to nasazení řešení do provozu.

4.4.1. Publikace dashboardu

Hotové řešení dashboardu bude umístěno na předem připravené místo v Microsoft Office Sharepoint Serveru (MOSS), který je ve společnosti používán jako publikační médium.

V dalším kroku publikace dashboardu bude nezbytné specifikovat, kteří uživatelé budou mít přístupové práva k dashboardu. To znamená vytvořit skupiny s různými právy - od správců dashboardu, kteří budou mít nejvyšší práva, až po běžné uživatele pouze s právem přístupu a zobrazení dashboardu.

Uživatelé budou přistupovat k dashboardu prostřednictvím intranetových stránek, kde bude nutná jejich autentifikace a ověření systémem.

4.4.2. Testování

Poté, až bude dashboard publikován, bude následovat jeho testování, které bude probíhat ve dvou krocích.

Prvním krokem bude kontrola, zda dashboard odpovídá požadavkům, které si stanovili uživatelé na začátku projektu ve fázi sběru uživatelských požadavků a zda je ovládání uživatelsky přívětivé.

Druhým krokem bude samotná kontrola dat, která budou zobrazována v jednotlivých reportech dashboardu. Kontrola bude probíhat následovně: uživatelé si vyexportují data ze zdrojových systémů a budou porovnávat, zda se uvedené hodnoty shodují s hodnotami v reportech. To znamená, bude použito výchozí metody analýzy dat, která byla až do současné doby ve společnosti používána a řešení využívající Business Intelligence.

Po důkladném otestování dashboardu bude následovat poslední krok, prováděný v rámci nasazení řešení do provozu, a to proces akceptace.

4.4.3. Proces akceptace

Po testovací fázi, v procesu nasazení řešení do provozu, bude s uživateli opět proveden skupinový rozhovor, kde bude zhodnocen celý projekt a bude provedena akceptace řešení.

Následně bude dashboard přesunut z testovacího Microsoft Office Sharepoint Serveru na produkční Microsoft Office Sharepoint Server a celý návrh se tak ukončí a uzavře.

Výsledné zobrazení řešení dashboardu je v této práci umístěno jako Příloha 2.

4.5. Shrnutí

V aplikační části je zachycen celý životní cyklus, který byl proveden při návrhu dashboardu.

Na začátku bylo nutné identifikovat a analyzovat uživatelské požadavky. Poté na základě nich navrhnout datový model datového skladu, dále způsob, jakým budou čerpána data ze zdrojových systémů (ETL proces) a následně navrhnout a vytvořit OLAP databázi. Po těchto krocích následoval návrh reportů, které budou obsahem dashboardu. Nakonec bylo popsáno nasazení řešení do provozu, od publikace, přes testování dashboardu, až k procesu akceptace celého řešení.

5. Závěr

Cílem práce bylo navrhnout řešení pro prezentaci prodejních výsledků společnosti formou dashboardu prostřednictvím nástrojů Microsoft Business Intelligence. Součástí návrhu bylo také specifikovat návrh způsobu čerpání dat ze zdrojových systémů, neboli ETL proces, a způsobu jejich uložení pro potřeby vizualizace pomocí navrhovaného dashboardu.

V práci bylo využito metody skupinového rozhovoru pro sběr požadavků na navrhovaný dashboard a dále pak bylo využito dimenzionálního modelování, kterým byl vytvořen datový model datového skladu, nad kterým byl poté vytvořen výsledný dashboard.

V první části této práce byla čtenáři přiblížena základní teorie a terminologie užívaná v oblasti Business Intelligence, která poté byla použita i v dalších částech práce.

Druhá část práce byla zaměřena na přiblížení výchozího stavu řešení prodejních výsledků, kde byla popsána (v té době) současná podoba řešení.

V poslední části práce byly specifikovány jednotlivé kroky, které byly nezbytné při návrhu řešení dashboardu pro prezentaci prodejních výsledků a manažerských informací společnosti.

V oblasti Business Intelligence je mnoho námětů k dalšímu řešení. Na tuto práci by mohlo například navazovat řešení pro prezentaci prodejních výsledků společnosti založené na mobilních platformách, které se stávají v současné době čím dál tím populárnější. Neustále totiž roste počet uživatelů tzv. chytrých telefonů a PDA zařízení a tak by se Business Intelligence mohlo ještě více přiblížit ke každému z nich.

Seznam použité literatury

BALLARD, Chuck. *Dimensional Modeling: In a Business Intelligence Environment* [online]. [20. 3. 2012]. Dostupné z: <http://www.redbooks.ibm.com/abstracts/sg247138.html>

ČVUT. *Datové kostky* [online]. [20. 3. 2012]. Dostupné z: <http://czm.fel.cvut.cz/vyuka/A4M33CPM/materialy.aspx>

ECKERSON, Wayne W. *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. 2. vyd. New Jersey: John Wiley&Sons, 2011. 318 s. ISBN 978-0-470-58983-0.

FEW, Stephen. *Information Dashboard Design: The Effective Visual Communication of Data*. Sebastopol: O'Reilly, 2006. 208 s. ISBN 978-0-596-10016-2.

GOLFARELLI, Matteo a Stefano RIZZI. *Data Warehouse Design: Modern Principles and Methodologies*. USA: McGraw-Hill books, 2009. 458 s. ISBN 978-0-07-161039-1.

HERYNEK, Josef a Štěpán KRÁSA a Martin MÁŠILKO. *Interaktivní dashboardy Versino Report Factory* [online]. [22. 3. 2012]. Dostupné z: <http://www.systemonline.cz/business-intelligence/interaktivni-dashboardy-versino-report-factory.htm>

INMON, William H. *Building the Data Warehouse*. 4. vyd. Indianapolis: John Wiley&Sons, 2005. 543 s. ISBN 978-0-7645-9944-6.

KAY, Russel. *Zaostřeno na datové kostky* [online]. [20. 3. 2012]. Dostupné z: <http://scienceworld.cz/technologie/zaostreno-na-datove-kostky-2223>

KIMBALL Ralph a Margy ROSS. *The Data Warehouse Toolkit: the Complete Guide to Dimensional Modeling*. 2. vyd. New York: John Wiley&Sons, 2002. 436 s. ISBN 0-471-20024-7.

LACKO, Ľuboslav. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby*. Brno: Computer Press, 2009. 456 s. ISBN 978-80-251-2887-9.

NOVOTNÝ, Ota a Jan POUR a David SLÁNSKÝ. *Business Intelligence: Jak využít bohatství ve vašich datech*. Praha: Grada Publishing, 2005. 256 s. ISBN 80 - 247- 1094 -3.

SCHEPS, Swain. *Business Intelligence For Dummies*. Indianapolis: Wiley Publishing, Inc., 2008. 358 s. ISBN 978-0-470-12723-0.

TVRDÍKOVÁ, Milena. *Aplikace moderních informačních technologií v řízení firmy: Nástroje ke zvyšování kvality informačních systémů*. Praha: Grada Publishing, 2008. 176 s. ISBN 978-80-247-2728-8.

ZÁDOVÁ, Vladimíra. *Multidimenzionální modelování v rámci analýzy a návrhu IS/ICT* [online]. [20. 3. 2012]. Dostupné z: <http://www.cssi.cz/cssi/multidimenzionalni-modelovani-v-ramci-analyzy-navrhu-ict>

Seznam zkratek

BI	- Business Intelligence
CRM	- Customer Relationship Management
CSV	- Comma-separated values
DMA	- Data Mart, Datové tržiště
DOLAP	- Desktop OLAP
DSA	- Data Staging Area, Dočasné uložení dat
DSS	- Decision Support Systems, Systémy pro podporu rozhodování řízení
DWH, DW	- Data Warehouse, Datový sklad
EAI	- Enterprise Application Integration, Integrace podnikových aplikací
EIS	- Executive Information Systems, Informační systémy pro vrcholové
ERP	- Enterprise Resource Planning
ETL	- Extract, Transformation, Load
HOLAP	- Hybrid OLAP
MIS	- Management Information Systems, Manažerské informační systémy
MOLAP	- Multidimensional OLAP
MOSS	- Microsoft Office Sharepoint Server
ODS	- Operational Data Store, Operativní uložení dat
OLAP	- On-Line Analytical Processing
OLTP	- On-Line Transaction Processing
ROLAP	- Relational OLAP
SCM	- Supply Chain Management
SSAS	- SQL Server Analysis Services
SSIS	- SQL Server Integration Services
SSRS	- SQL Server Reporting Services

Prohlášení o využití výsledků diplomové práce

Prohlašuji, že,

- jsem byla seznámena s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. - autorský zákon, zejména § 35 - užití díla v rámci občanských a náboženských obřadů, v rámci školních představení a užití díla školního a § 60 - školní dílo;
- beru na vědomí, že Vysoká škola báňská - Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užít (§ 35 odst. 3);
- souhlasím s tím, že diplomová práce bude v elektronické podobě archivována v Ústřední knihovně VŠB-TUO a jeden výtisk bude uložen u vedoucího diplomové práce. Souhlasím s tím, že bibliografické údaje o diplomové práci budou zveřejněny v informačním systému VŠB-TUO;
- bylo sjednáno, že s VŠB-TUO, v případě zájmu z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona;
- bylo sjednáno, že užít své dílo, diplomovou práci, nebo poskytnout licenci k jejímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

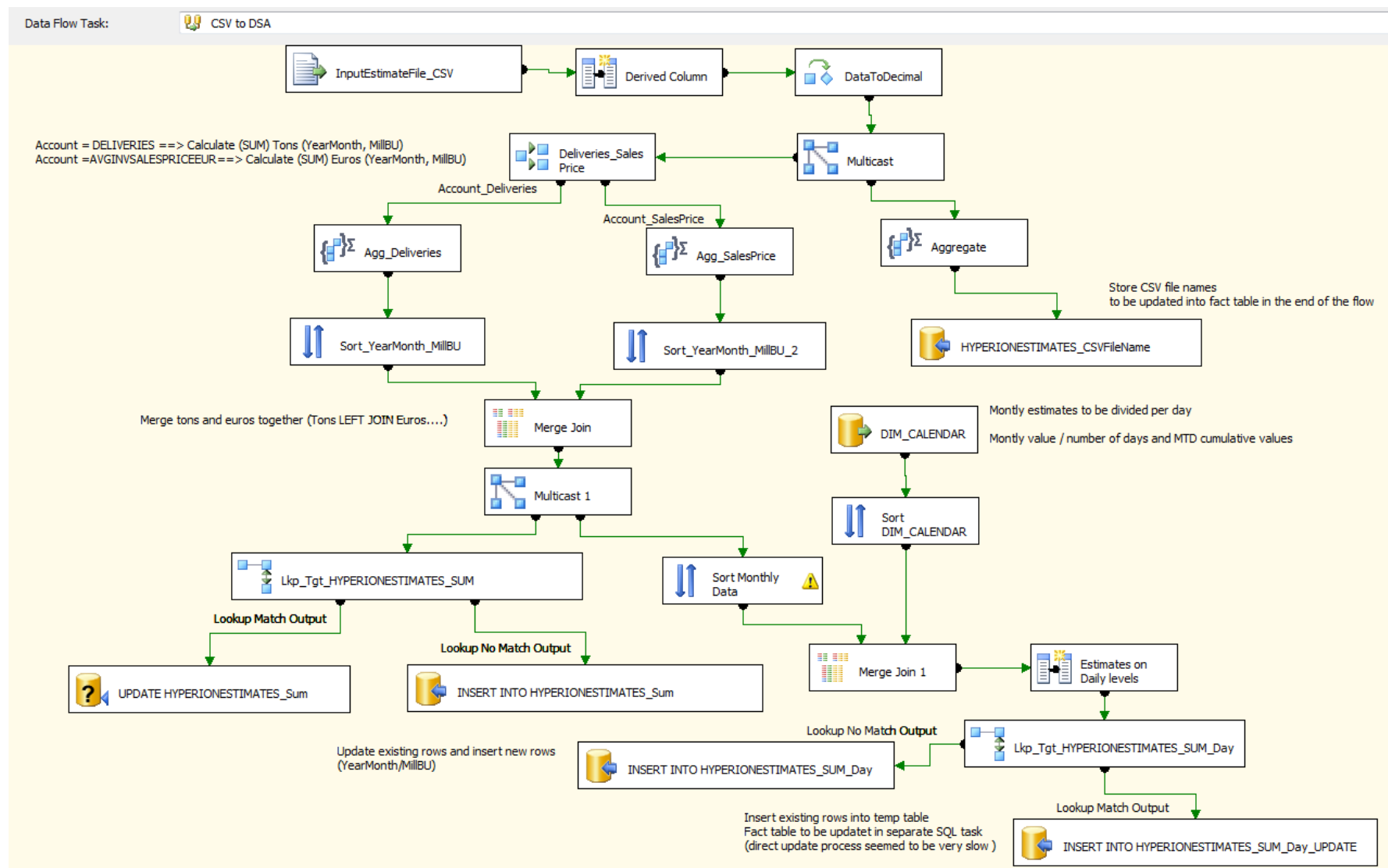
V Ostravě dne 27. 4. 2012

.....
Jana Kaděrová

Seznam příloh

Příloha 1 - ETL proces (pokračování).....	1
Příloha 2 - Dashboard pro prezentaci prodejních výsledků společnosti.....	2

Přílohy



Příloha 1 - ETL proces (pokračování)

[Zdroj: vlastní]



Příloha 2 - Dashboard pro prezentaci prodejních výsledků společnosti

[Zdroj: vlastní]